Gefitinib Sensitivity-Related Gene Expression and Products and Methods Related Thereto

Field of the Invention

This invention generally relates to methods to screen for patients that are predicted to benefit from therapeutic administration of gefitinib, as well as methods to identify compounds that interact with the epidermal growth factor receptor (EGFR) pathway to allow or enhance responsiveness to EGFR inhibitors, and products and methods related thereto.

Background of the Invention

Lung Cancer is the leading cause of death from cancer worldwide. Chemotherapy is the mainstay of treatment for lung cancer. However, less than a third of patients with advanced stages of non-small cell lung cancer (NSCLC) respond to the best two chemotherapy drug combinations. Therefore, novel agents that target cancer specific biological pathways are needed.

The epidermal growth factor receptor (EGFR) is one of the most appealing targets for novel therapies for cancer. EGFR plays a major role in transmitting stimuli that lead to proliferation, growth and survival of various cancer types, including, but not limited to, NSCLC. Ligand binding to the EGFR receptor leads to homo- or heterodimerization of EGFR with other ErbB receptors. EGFR is overexpressed in a large proportion of invasive NSCLC and in premalignant bronchial lesions. Bronchioloalveolar carcinoma (BAC), a subtype of non-small cell lung cancer, represents the major form of lung cancer in non-smoking females and is rising in frequency, and epidermal growth factor receptor (EGFR) is expressed with high frequency in BAC. Unfortunately, the response of BACs to conventional chemotherapy is poor. Activation of EGFR leads to simultaneous activation of several signaling cascades including the MAPK pathway, the protein kinase C (PKC) pathway and the PI(3)K-activated AKT pathway (Fig. 1). EGFR signaling translated in the nucleus leads to cancer cell proliferation and survival.

Targeted therapy against the EGFR receptor has produced response rates of 25-30% as first line treatment and 11-20% in 2nd and 3rd line settings (e.g., chemo-refractory advanced stage NSCLC). For example, in phase II clinical trials, 11-20% of patients with chemo-refractory advanced stage NSCLC responded to treatment with the EGFR tyrosine kinase inhibitor gefitinib (commercially available as Iressa®, ZD1839). A trial

evaluating the activity of the EGFR inhibitor, erlotinib (Tarceva®, OSI-774) has been completed and the results will be reported in the near future. A retrospective analysis of 140 patients responding to treatment with gefitinib revealed that the presence of BAC features ($p=0.005$) and being a never smoker ($p=0.007$) were the only independent predictors of response to gefitinib. These data suggest that EGFR inhibitor therapy is more active in BAC and in non-smokers.

However, currently, there are no selection criteria for determining which NSCLC patients will benefit from treatment with EGFR inhibitors such as gefitinib. Moreover, EGFR expression does not predict gefitinib sensitivity. Therefore, despite the correlation of tumor histology and smoking history with gefitinib response, it is of great importance to identify molecular molecules that influence gefitinib responsiveness, and to develop adjuvant treatments that enhance the response. To accomplish this goal, there is a need in the art to define critical aspects of EGFR signaling and to identify which molecules interact with the EGFR pathway to dictate responsiveness to EGFR inhibitors.

                              Summary of the Invention

One embodiment of the present invention relates to a method to select a cancer patient who is predicted to benefit from therapeutic administration of an EGFR inhibitor, an agonist thereof, or a drug having substantially similar biological activity as EGFR inhibitor. The method includes the steps of: (a) providing a sample of tumor cells from a patient to be tested; (b) detecting in the sample the expression of one or more genes chosen from a panel of genes whose expression has been correlated with sensitivity or resistance to an EGFR inhibitor; (c) comparing the level of expression of the gene or genes detected in the patient sample to a level of expression of the gene or genes that has been correlated with sensitivity or resistance to the EGFR inhibitor; and (d) selecting the patient as being predicted to benefit from therapeutic administration of the EGFR inhibitor, if the expression of the gene or genes in the patient's tumor cells is statistically more similar to the expression levels of the gene or genes that has been correlated with sensitivity to the EGFR inhibitor than to resistance to the EGFR inhibitor.

In one aspect, the panel of genes in (b) is identified by a method comprising: (a) providing a sample of cells that are sensitive or resistant to treatment with the EGFR inhibitor; (b) detecting the expression of at least one gene in the EGFR inhibitor-sensitive cells as compared to the level of expression of the gene or genes in the EGFR inhibitor-resistant cells; and (c) identifying a gene or genes having a level of expression in EGFR

inhibitor-sensitive cells that is statistically significantly different than the level of expression of the gene or genes in EGFR inhibitor-resistant cells, as potentially being a molecule that interacts with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors.

In another aspect, the EGFR inhibitor is gefitinib. In this aspect, step (b) can include, in one embodiment, detecting in the sample the expression of one or more genes chosen from a gene comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194. Step (c) comprises comparing the level of expression of the gene or genes detected in the patient sample to a level of expression of the gene or genes that has been correlated with sensitivity or resistance to gefitinib. Step (d) comprises selecting the patient as being predicted to benefit from therapeutic administration of gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib, if the expression of the gene or genes in the patient's tumor cells is statistically more similar to the expression levels of the gene or genes that has been correlated with sensitivity to gefitinib than to resistance to gefitinib.

In any of the embodiments above, the method can include detecting expression of at least two genes in (b), at least three genes in (b), at least four genes in (b), at least five genes in (b), at least 10 genes in (b), at least 25 genes in (b), at least 50 genes from in (b), at least 100 genes in (b), at least 150 genes in (b), or up to all of the genes in the panel of genes.

In one aspect of this method, expression of the gene or genes is detected by measuring amounts of transcripts of the gene in the tumor cells. In another aspect, expression of the gene or genes is detected by detecting hybridization of at least a portion of the gene or a transcript thereof to a nucleic acid molecule comprising a portion of the gene or a transcript thereof in a nucleic acid array. In another aspect, expression of the gene is detected by detecting the production of a protein encoded by the gene. In yet another aspect, the method includes detecting expression of at least one gene selected from the group consisting of: E-cadherin (represented by SEQ ID NO:3) and ErbB3 (represented by SEQ ID NO:15 or SEQ ID NO:133). For example, the method can include detecting expression of at least one gene selected from the group consisting of ZEB1 and SIP1.

In one aspect of this method, the method includes comparing the expression of the gene or genes to expression of the gene or genes in a cell from a non-cancerous cell of the same type. In another aspect, the method includes comparing the expression of the gene or genes to expression of the gene or genes in an autologous, non-cancerous cell from the patient. In another aspect, the method includes comparing the expression of the gene or genes to expression of the gene or genes in a control cell that is resistant to the EGFR inhibitor. In yet another aspect, the method includes comparing the expression of the gene or genes to expression of the gene or genes in a control cell that is sensitive to the EGFR inhibitor. In another aspect, control expression levels of the gene or genes that has been correlated with sensitivity and/or resistance to the EGFR inhibitor has been predetermined.

Yet another embodiment of the present invention relates to a method to identify molecules that interact with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors. The method includes the steps of: (a) providing a sample of cells that are sensitive or resistant to treatment with gefitinib; (b) detecting the expression of at least one gene in the gefitinib-sensitive cells as compared to the level of expression of the gene or genes in the gefitinib-resistant cells; and (c) identifying a gene or genes having a level of expression in gefitinib-sensitive cells that is statistically significantly different than the level of expression of the gene or genes in gefitinib-resistant cells, as potentially being a molecule that interacts with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors.

Another embodiment of the present invention relates to a plurality of polynucleotides for the detection of the expression of genes that are indicative of sensitivity or resistance to gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib. The plurality of polynucleotides consists of at least two polynucleotides, wherein each polynucleotide is at least 5 nucleotides in length, and wherein each polynucleotide is complementary to an RNA transcript, or nucleotide derived therefrom, of a gene that is regulated differently in gefitinib-sensitive tumor cells as compared to gefitinib-resistant cells. In one aspect, each polynucleotide is complementary to an RNA transcript, or a polynucleotide derived therefrom, of a gene comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194. In another aspect, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript,

or a nucleotide derived therefrom, of at least two genes comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194. In another aspect, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least five genes, at least 10 genes, at least 25 genes, at least 50 genes, at least 100 genes, at least 150 genes, or up to all of the genes, comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194. In one aspect, the polynucleotide probes are immobilized on a substrate. In another aspect, the polynucleotide probes are hybridizable array elements in a microarray. In yet another aspect, the polynucleotide probes are conjugated to detectable markers.

Yet another embodiment of the present invention relates to a plurality of antibodies, antigen binding fragments thereof, or antigen binding peptides, for the detection of the expression of genes that are indicative of sensitivity or resistance to gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib. The plurality of antibodies, antigen binding fragments thereof, or antigen binding peptides consists of at least two antibodies, antigen binding fragments thereof, or antigen binding peptides, each of which selectively binds to a protein encoded by a gene comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194.

Another embodiment of the present invention relates to a method to identify a compound with the potential to enhance the efficacy of EGFR inhibitors. The method includes the steps of: (a) contacting a test compound with a cell that expresses at least one gene, wherein said gene is selected from any one of the genes comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194; (b) identifying compounds selected from the group consisting of: (i) compounds that increase the expression or activity of the gene or genes in (a), or the proteins encoded thereby, that are correlated with sensitivity to gefitinib; and (ii) compounds that decrease the expression or activity of genes in (a), or the proteins encoded thereby, that are correlated with resistance to gefitinib. The compounds are identified as having the potential to enhance the efficacy of EGFR inhibitors. In one aspect of this embodiment, the cell expresses a gene encoding E-cadherin or ErbB3, and wherein step (b) comprises identifying compounds that increase the expression or activity

of E-cadherin or ErbB3 or the gene encoding E-cadherin or ErbB3. In another aspect of this embodiment, the cell expresses a gene encoding ZEB1 and SIP1, wherein step (b) comprises identifying compounds that decrease the expression or activity ZEB1 or SIP1 or the gene encoding ZEB1 or SIP1.

Another embodiment of the present invention relates to a method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound identified by the method described above.

Yet another embodiment of the present invention relates to a method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound that upregulates the expression or activity of E-cadherin or ErbB3 or the gene encoding E-cadherin or ErbB3 in the tumor cells of the patient.

Another embodiment of the present invention relates to a method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound that downregulates the expression of ZEB1 or SIP1 or the gene encoding ZEB1 or SIP1 in the tumor cells of the patient.

Brief Description of the Figures of the Invention

Fig. 1 is a schematic diagram showing the activation of signaling cascades from EGFR.

Fig. 2 is a schematic diagram showing E-cadherin regulation.

Fig. 3 is a digital image showing the expression of EGFR and phosphorylated EGFR in NSCLC cell lines.

Fig. 4 is a digital image showing that ZD1839 downregulates pEGFR in sensitive NSCLC cell lines.

Fig. 5 is a line graph showing the effects of gefitinib on A549 NSCLC xenografts.

Fig. 6 is a bar graph showing the expression of E-cadherin in NSCLC cell lines using GeneSpring analysis of microarrays.

Fig. 7 is a digital image showing Western blot analysis of E-cadherin expression in NSCLC cell lines.

Fig. 8 is a bar graph showing real time RT-PCR analysis of ZEB1 and SIP1 expression in NSCLC cell lines.

Fig. 9 is a schematic drawing showing the use of siRNA to silence the E-cadherin transcriptional repressors, SIP1 and ZEB1 to determine the effect on NSCLC cell line responses to ZD1839.

Detailed Description of the Invention

The present invention generally relates to the identification, provision and use of a panel of biomarkers that predict sensitivity or resistance to gefitinib and other EGFR inhibitors, and products and processes related thereto. Specifically, the present inventors have used NSCLC cell lines with varying sensitivity to the EGFR inhibitor, gefitinib, to define the novel panel of biomarkers as described herein. In order to identify a marker panel that could be used for selection of NSCLC patients who will respond to gefitinib treatment, the inventors undertook preclinical *in vitro* studies using NSCLC cell lines. Based on the therapeutic response to gefitinib by using the IC_{50} definition (i.e., the concentration of agent needed to kill 50% of the tumor cells in a cell culture), the present inventors have classified the cell lines as sensitive ($IC_{50} < 1$ μM), resistant ($IC_{50} > 10$ μM), or having intermediate sensitivity (1 μM $< IC_{50} < 10$ μM) to gefitinib. The cell lines were characterized by gene microarray analysis (Affymetrix™ microarray Human Genome U133 set, 39,000 genes). By comparing the gene microarray results from sensitive and resistant cell lines, the inventors have identified a panel of genes that can discriminate between sensitive and resistant cell lines. These biomarkers (i.e., the genes identified) will be of great clinical significance in selecting NSCLC patients/human tumors which will respond to this agent. The biomarkers identified by the present invention, and their expression levels in gefitinib sensitive and resistant cells, are listed in Table 1, and the nucleotide sequences representing such biomarkers are represented herein by SEQ ID NOs:1-194. The nucleic acid sequences represented by SEQ ID NOs:1-194 include transcripts or nucleotides derived therefrom (e.g., cDNA) expressed by the gene biomarkers in Table 1. It is to be understood that the present invention expressly covers additional genes that can be elucidated using substantially the same techniques used to identify the genes in Table 1 and that any of such additional genes can be used in the methods and products described herein for the genes and probe sets in Table 1. Any reference to database Accession numbers or other information regarding the genes and probe sets in Table 1 is hereby incorporated by reference in its entirety. For each biomarker listed in Table 1, the following information is provided: (1) the probe set ID number given by Affymetrix™ for the set of features on the array representing the indicated gene; (2) the parametric p-value, indicating the statistical significance of that individual gene expression difference; (3) the mean intensity of expression of each gene in a gefitinib-sensitive and a gefitinib-resistant cell line; (4) the HUGO-approved symbol

for the gene, where one exists; (5) the sequence identifier representing a nucleotide sequence found in or transcribed by the gene; and (6) the name or title of the gene, where one is given. It is noted that sometimes two probe sets in Table 1 will refer to a single gene, and these duplications have been maintained because they are believed to reflect different splice variants of that gene. In such a case, the associated sequence files will reflect the different splicotypes for that gene. The genes in Table 1 have been sorted by their parametric p-value to indicate the genes that are most highly regulated by gefitinib first.

In addition, the present invention will also be useful for the validation in other studies of the clinical significance of many of the specific biomarkers described herein, as well as the identification of preferred biomarker profiles, highly sensitive biomarkers, and targets for the design of novel therapeutic products and strategies. The biomarkers described herein are particularly useful in clinical practice to select the patients who will benefit most from EGFR inhibitor treatment and in specific embodiments, from gefitinib treatment.

The present inventors have already used the biomarkers described herein to identify specific targets for the further development of diagnostic and therapeutic approaches used in cancer, and these studies are described in detail in the Examples. For example, E-cadherin is a calcium-dependent epithelial cell adhesion molecule that plays an important role in tumor invasiveness and metastatic potential. Reduced E-cadherin expression is associated with tumor cell dedifferentiation, advanced stage and reduced survival in patients with NSCLC. Using Western blot analysis, E-cadherin was expressed in three cell lines highly sensitive to gefitinib and its expression was lacking in six gefitinib resistant cell lines tested. Real-time RT-PCR was used to evaluate the gene expression pattern in 11 NSCLC cell lines and compared to gene expression in normal bronchial epithelium. E-cadherin expression was elevated in cell lines sensitive to gefitinib and downregulated in the resistant cell lines as compared to the normal bronchial epithelium. The expression of E-cadherin is regulated by zinc finger inhibitory proteins by the recruitment of histone deacetylases (HDAC). Using real-time RT-PCR, the expression of the two zinc-finger transcription factors, δEF1/ZEB1 and SIP1/ZEB2, involved in E-cadherin repression was evaluated. Results showed that ZEB1 was expressed in gefitinib resistant cell lines and its expression was lacking in gefitinib sensitive cell lines. The present inventors have also found that δEF1/ZEB1 and

SIP1/ZEB2 may regulate Her3, which is an EGFR heterodimer. These data indicate that the expression of ZEB1 may predict resistance to EGFR tyrosine kinase inhibitors and future studies directed at modulating the regulation of E-cadherin expression are expected to enhance the activity of EGFR inhibitors in NSCLC.

Finally, the present invention also relates to protein profiles which can discriminate between sensitive and resistant NSCLC tumors.

Prior to the present invention, to the best of the present inventors' knowledge, no single marker, or marker panel, has been demonstrated to be useful for selection of lung cancer patients who will benefit from EGFR inhibitors, and particularly, gefitinib, treatment. Nor are there any such markers (related to EGFR inhibitors) identified for other types of cancer.

Accordingly, using the gene expression profiles disclosed in Table 1 for gefitinib-sensitive and -resistant cells, one can rapidly, effectively and efficiently screen patients/human tumors for a level of sensitivity or resistance to gefitinib and also to other EGFR inhibitors having biological activity substantially similar to gefitinib (i.e., drugs having similar activities, gefitinib agonists and other derivatives). The results will allow for the identification of tumors/patients that are likely to benefit from administration of the drug and therefore, the genes are used to enhance the ability of the clinician to develop prognosis and treatment protocols for the individual patient. In addition, genes identified in Table 1 can be further validated as targets and then used in assays to identify therapeutic reagents useful for regulating the expression or activity of the target in a manner that improves sensitivity of a cell to gefitinib or analogs thereof. The knowledge provided from the expression profile of genes described herein and the identification additional genes using similar methods can also be used to identify the molecular mechanisms of EGFR inhibition, such knowledge being useful for the further development of new therapies and even analogs of gefitinib or other EGFR inhibitors with improved efficacies in cancer treatment. Moreover, given the knowledge of these genes, one can produce novel combinations of polynucleotides and/or antibodies and/or peptides for use in the various assays, diagnostic and/or therapeutic approaches described herein.

Finally, the present invention is also illustrative of methods by which patients can be evaluated for predicted sensitivity or resistance to EGFR inhibitors other than gefitinib, and of methods of identifying additional genes and gene panels that are

regulated differentially by cells that are sensitive to or resistant to gefitinib or other EGFR inhibitors. Such genes and panels of genes can then be used in the assays and methods described herein and as targets useful for the development of novel EGFR inhibitors and therapeutic formulations.

5   Various definitions and aspects of the invention will be described below, but the invention is not limited to any specific embodiments that may be used for illustrative or exemplary purposes.

According to the present invention, in general, the biological activity or biological action of a protein refers to any function(s) exhibited or performed by the protein that is 10   ascribed to the naturally occurring form of the protein as measured or observed *in vivo* (i.e., in the natural physiological environment of the protein) or *in vitro* (i.e., under laboratory conditions). Modifications of a protein, such as in a homologue or mimetic (discussed below), may result in proteins having the same biological activity as the naturally occurring protein, or in proteins having decreased or increased biological 15   activity as compared to the naturally occurring protein. Modifications which result in a decrease in protein expression or a decrease in the activity of the protein, can be referred to as inactivation (complete or partial), down-regulation, or decreased action of a protein. Similarly, modifications which result in an increase in protein expression or an increase in the activity of the protein, can be referred to as amplification, overproduction, activation, 20   enhancement, up-regulation or increased action of a protein.

According to the present invention, a "downstream gene" or "endpoint gene" is any gene, the expression of which is regulated (up or down) within a gefitinib sensitive or resistant cell. Selected sets of one, two, and preferably several or many of the genes (up to the number equivalent to all of the genes) of this invention can be used as end-points 25   for rapid screening of patient cells for sensitivity or resistance to EGFR inhibitors such as gefitinib and for the other methods as described herein, including the identification of novel targets for the development of new cancer therapeutics.

As used herein, the term "homologue" is used to refer to a protein or peptide which differs from a naturally occurring protein or peptide (i.e., the "prototype" or "wild- 30   type" protein) by minor modifications to the naturally occurring protein or peptide, but which maintains the basic protein and side chain structure of the naturally occurring form. Such changes include, but are not limited to: changes in one or a few amino acid side chains; changes one or a few amino acids, including deletions (e.g., a truncated version of

the protein or peptide) insertions and/or substitutions; changes in stereochemistry of one or a few atoms; and/or minor derivatizations, including but not limited to: methylation, glycosylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, amidation and/or addition of glycosylphosphatidyl inositol. A homologue can have either enhanced, decreased, or substantially similar properties as compared to the naturally occurring protein or peptide. A homologue can include an agonist of a protein or an antagonist of a protein.

Homologues can be the result of natural allelic variation or natural mutation. A naturally occurring allelic variant of a nucleic acid encoding a protein is a gene that occurs at essentially the same locus (or loci) in the genome as the gene which encodes such protein, but which, due to natural variations caused by, for example, mutation or recombination, has a similar but not identical sequence. Allelic variants typically encode proteins having similar activity to that of the protein encoded by the gene to which they are being compared. One class of allelic variants can encode the same protein but have different nucleic acid sequences due to the degeneracy of the genetic code. Allelic variants can also comprise alterations in the 5' or 3' untranslated regions of the gene (e.g., in regulatory control regions). Allelic variants are well known to those skilled in the art.

An agonist can be any compound which is capable of mimicking, duplicating or approximating the biological activity of a naturally occurring or specified protein, for example, by associating with (e.g., binding to) or activating a protein (e.g., a receptor) to which the natural protein binds, so that activity that would be produced with the natural protein is stimulated, induced, increased, or enhanced. For example, an agonist can include, but is not limited to, a protein, compound, or an antibody that selectively binds to and activates or increases the activation of a receptor bound by the natural protein, other homologues of the natural protein, and any suitable product of drug design that is characterized by its ability to agonize (e.g., stimulate, induce, increase, enhance) the biological activity of a naturally occurring protein.

An antagonist refers to any compound or agent which is capable of acting in a manner that is antagonistic to (e.g., against, a reversal of, contrary to) the action of the natural agonist, for example by interacting with another protein or molecule in a manner that the biological activity of the naturally occurring protein or agonist is decreased (e.g., reduced, inhibited, blocked). Such a compound can include, but is not limited to, an antibody that selectively binds to and blocks access to a protein by its natural ligand, or

reduces or inhibits the activity of a protein, a product of drug design that blocks the protein or reduces the biological activity of the protein, an anti-sense nucleic acid molecule that binds to a nucleic acid molecule encoding the protein and prevents expression of the protein, a ribozyme that binds to the RNA and prevents expression of the protein, RNAi, an aptamer, and a soluble protein, which competes with a natural receptor or ligand.

Agonists and antagonists that are products of drug design can be produced using various methods known in the art. Various methods of drug design, useful to design mimetics or other compounds useful in the present invention are disclosed in Maulik et al., 1997, *Molecular Biotechnology: Therapeutic Applications and Strategies*, Wiley-Liss, Inc., which is incorporated herein by reference in its entirety. An agonist or antagonist can be obtained, for example, from molecular diversity strategies (a combination of related strategies allowing the rapid construction of large, chemically diverse molecule libraries), libraries of natural or synthetic compounds, in particular from chemical or combinatorial libraries (i.e., libraries of compounds that differ in sequence or size but that have the similar building blocks) or by rational, directed or random drug design. See for example, Maulik et al., *supra.*

In a molecular diversity strategy, large compound libraries are synthesized, for example, from peptides, oligonucleotides, natural or synthetic steroidal compounds, carbohydrates and/or natural or synthetic organic and non-steroidal molecules, using biological, enzymatic and/or chemical approaches. The critical parameters in developing a molecular diversity strategy include subunit diversity, molecular size, and library diversity. The general goal of screening such libraries is to utilize sequential application of combinatorial selection to obtain high-affinity ligands for a desired target, and then to optimize the lead molecules by either random or directed design strategies. Methods of molecular diversity are described in detail in Maulik, et al., *ibid.*

As used herein, the term "mimetic" is used to refer to any natural or synthetic compound, peptide, oligonucleotide, carbohydrate and/or natural or synthetic organic molecule that is able to mimic the biological action of a naturally occurring or known synthetic compound.

As used herein, the term "putative regulatory compound" or "putative regulatory ligand" refers to compounds having an unknown regulatory activity, at least with respect to the ability of such compounds to regulate the expression or biological activity of a gene

or protein encoded thereby, or to regulate sensitivity or resistance to an EGFR inhibitor as encompassed by the present invention.

In accordance with the present invention, an isolated polynucleotide, which phrase can be used interchangeably with "an isolated nucleic acid molecule", is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. Polynucleotides useful in the plurality of polynucleotides of the present invention (described below) are typically a portion of a gene or transcript thereof of the present invention that is suitable for use, for example, as a hybridization probe or PCR primer for the identification of a full-length gene, a transcript thereof, or a polynucleotide derived from the gene or transcript (e.g., cDNA), in a given sample (e.g., a cell sample). An isolated nucleic acid molecule can include a gene or a portion of a gene (e.g., the regulatory region or promoter), for example, to produce a reporter construct according to the present invention. An isolated nucleic acid molecule that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes naturally found on the same chromosome. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecules can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Although the phrase "nucleic acid molecule" or "polynucleotide" primarily refers to the physical nucleic acid molecule and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the nucleic acid molecule, the two phrases can be used interchangeably, especially with respect to a nucleic acid molecule, or a nucleic acid sequence, being capable of encoding a protein.

Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. Isolated nucleic acid molecules include natural nucleic acid molecules and homologues thereof, including, but not limited to,

natural allelic variants and modified nucleic acid molecules in which nucleotides have been inserted, deleted, substituted, and/or inverted in such a manner that such modifications provide the desired effect on the biological activity of the protein as described herein. Protein homologues (e.g., proteins encoded by nucleic acid homologues) have been discussed in detail above.

The minimum size of a nucleic acid molecule or polynucleotide of the present invention is a size sufficient to encode a protein having a desired biological activity, sufficient to form a probe or oligonucleotide primer that is capable of forming a stable hybrid with the complementary sequence of a nucleic acid molecule encoding the natural protein (e.g., under moderate, high or very high stringency conditions), or to otherwise be used as a target in an assay or in any therapeutic method discussed herein. If the polynucleotide is an oligonucleotide probe or primer, the size of the polynucleotide can be dependent on nucleic acid composition and percent homology or identity between the nucleic acid molecule and a complementary sequence as well as upon hybridization conditions *per se* (e.g., temperature, salt concentration, and formamide concentration). The minimum size of a polynucleotide that is used as an oligonucleotide probe or primer is at least about 5 nucleotides in length, and preferably ranges from about 5 to about 50 or about 500 nucleotides, including any length in between, in whole number increments (i.e., 5, 6, 7, 8, 9, 10,...33, 34,...256, 257,...500), and more preferably from about 10 to about 40 nucleotides, and most preferably from about 15 to about 40 nucleotides in length. In one aspect, the oligonucleotide primer or probe is typically at least about 12 to about 15 nucleotides in length if the nucleic acid molecules are GC-rich and at least about 15 to about 18 bases in length if they are AT-rich. There is no limit, other than a practical limit, on the maximal size of a nucleic acid molecule of the present invention, in that the nucleic acid molecule can include a portion of a protein-encoding sequence or a nucleic acid sequence encoding a full-length protein.

An isolated protein, according to the present invention, is a protein (including a peptide) that has been removed from its natural milieu (i.e., that has been subject to human manipulation) and can include purified proteins, partially purified proteins, recombinantly produced proteins, and synthetically produced proteins, for example. As such, "isolated" does not reflect the extent to which the protein has been purified. An isolated protein useful as an antagonist or agonist according to the present invention can be isolated from its natural source, produced recombinantly or produced synthetically.

Smaller peptides useful as regulatory peptides are typically produced synthetically by methods well known to those of skill in the art.

According to the present invention, the phrase "selectively binds to" refers to the ability of an antibody, antigen binding fragment or binding partner (antigen binding peptide) to preferentially bind to specified proteins. More specifically, the phrase "selectively binds" refers to the specific binding of one protein to another (*e.g.*, an antibody, fragment thereof, or binding partner to an antigen), wherein the level of binding, as measured by any standard assay (*e.g.*, an immunoassay), is statistically significantly higher than the background control for the assay. For example, when performing an immunoassay, controls typically include a reaction well/tube that contain antibody or antigen binding fragment alone (*i.e.*, in the absence of antigen), wherein an amount of reactivity (*e.g.*, non-specific binding to the well) by the antibody or antigen binding fragment thereof in the absence of the antigen is considered to be background. Binding can be measured using a variety of methods standard in the art including enzyme immunoassays (*e.g.*, ELISA), immunoblot assays, etc.).

In some embodiments of the present invention, a compound is contacted with one or more nucleic acids or proteins. Such methods can include cell-based assays, or non-cell-based assay. In one embodiment, a target gene is expressed by a cell (i.e., a cell-based assay). In one embodiment, the conditions under which a cell expressing a target is contacted with a putative regulatory compound, such as by mixing, are conditions in which the expression or biological activity of the target (gene or protein encoded thereby) is not stimulated (activated) if essentially no regulatory compound is present. For example, such conditions include normal culture conditions in the absence of a known activating compound or other equivalent stimulus. The putative regulatory compound is then contacted with the cell. In this embodiment, the step of detecting is designed to indicate whether the putative regulatory compound alters the expression and/or biological activity of the gene or protein target as compared to in the absence of the putative regulatory compound (i.e., the background level).

In accordance with the present invention, a cell-based assay as described herein is conducted under conditions which are effective to screen for regulatory compounds or to profile gene expression as described in the methods of the present invention. Effective conditions include, but are not limited to, appropriate media, temperature, pH and oxygen conditions that permit the growth of the cell that expresses the receptor. An appropriate,

or effective, medium is typically a solid or liquid medium comprising growth factors and assimilable carbon, nitrogen and phosphate sources, as well as appropriate salts, minerals, metals and other nutrients, such as vitamins. Culturing is carried out at a temperature, pH and oxygen content appropriate for the cell. Such culturing conditions are within the

5   expertise of one of ordinary skill in the art.

Cells that are useful in the cell-based assays of the present invention include any cell that expresses a gene that is to be investigated as a target, or in the diagnostic assays described herein, any cell that is isolated from a patient, including normal or malignant (tumor) cells.

10   According to the present invention, the method includes the step of detecting the expression of at least one, and preferably more than one, and most preferably, several, of the genes that are regulated differently in EGFR inhibitor-sensitive versus EGFR inhibitor-resistant cells, and particularly, of the genes that have now been shown to be regulated differently in gefitinib-sensitive versus gefitinib-resistant cells, by the present

15   inventors. As used herein, the term "expression", when used in connection with detecting the expression of a gene, can refer to detecting transcription of the gene and/or to detecting translation of the gene. To detect expression of a gene refers to the act of actively determining whether a gene is expressed or not. This can include determining whether the gene expression is upregulated as compared to a control, downregulated as

20   compared to a control, or unchanged as compared to a control. Therefore, the step of detecting expression does not require that expression of the gene actually is upregulated or downregulated, but rather, can also include detecting that the expression of the gene has not changed (i.e., detecting no expression of the gene or no change in expression of the gene).

25   The present method includes the step of detecting the expression of at least one gene set forth in Table 1. In a preferred embodiment, the step of detecting includes detecting the expression of at least 2 genes, and preferably at least 3 genes, and more preferably at least 4 genes, and more preferably at least 5 genes, and more preferably at least 6 genes, and more preferably at least 7 genes, and more preferably at least 8 genes,

30   and more preferably at least 9 genes, and more preferably at least 10 genes, and more preferably at least 11 genes, and more preferably at least 12 genes, and more preferably at least 13 genes, and more preferably at least 14 genes, and more preferably at least 15 genes, and so on, in increments of one (i.e., 1, 2, 3,...12, 13,...56, 57,...78, 79...), up to

detecting expression of all of the genes disclosed herein in Table 1. For example, in one aspect of the invention, the expression of at least five genes is detected, and in another aspect, the expression of at least 10 genes is detected, and in another aspect, the expression of at least 25 genes is detected, and in another aspect, the expression of at least 50 genes is detected, and in another aspect, the expression of at least 100 genes is detected, and in another aspect, the expression of at least 150 genes is detected. Preferably, larger numbers of genes in Table 1 are detected, as this will increase the sensitivity of the detection method. Analysis of a number of genes greater than 1 can be accomplished simultaneously, sequentially, or cumulatively.

In another embodiment of the invention, detecting in the sample the expression of one or more genes chosen from a panel of genes whose expression has been correlated with sensitivity or resistance to an EGFR inhibitor. For example, such genes can be identified using the methods for identifying the genes whose expression is correlated with gefitinib-resistance or sensitivity as described herein. In one aspect, the panel of genes is identified by a method comprising: (a) providing a sample of cells that are sensitive or resistant to treatment with the EGFR inhibitor; (b) detecting the expression of at least one gene in the EGFR inhibitor-sensitive cells as compared to the level of expression of the gene or genes in the EGFR inhibitor-resistant cells; and (c) identifying a gene or genes having a level of expression in EGFR inhibitor-sensitive cells that is statistically significantly different than the level of expression of the gene or genes in EGFR inhibitor-resistant cells, as potentially being a molecule that interacts with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors. The present invention is not intended to be limited solely to the biomarkers listed in Table 1. Rather, the biomarkers of Table 1 illustrate various aspects of the invention that can now be achieved given the discoveries by the inventors. Therefore, although many of the embodiments below are discussed in terms gefitinib, it is to be understood that the methods of the invention can be extended to other EGFR inhibitors, and particularly to those that are similar in structure and/or function to gefitinib, including agonists of gefitinib.

The first steps of the method to select a cancer patient that is predicted to benefit from therapeutic administration of an EGFR inhibitor, an agonist thereof, or a drug having substantially similar biological activity as EGFR inhibitor of the present invention, includes providing a patient sample (also called a test sample) and detecting in the sample the expression of a gene or genes. Suitable methods of obtaining a patient sample are

18

known to a person of skill in the art. A patient sample can include any bodily fluid or tissue from a patient that may contain tumor cells or proteins of tumor cells. More specifically, according to the present invention, the term "test sample" or "patient sample" can be used generally to refer to a sample of any type which contains cells or products

5   that have been secreted from cells to be evaluated by the present method, including but not limited to, a sample of isolated cells, a tissue sample and/or a bodily fluid sample. According to the present invention, a sample of isolated cells is a specimen of cells, typically in suspension or separated from connective tissue which may have connected the cells within a tissue *in vivo*, which have been collected from an organ, tissue or fluid

10   by any suitable method which results in the collection of a suitable number of cells for evaluation by the method of the present invention. The cells in the cell sample are not necessarily of the same type, although purification methods can be used to enrich for the type of cells that are preferably evaluated. Cells can be obtained, for example, by scraping of a tissue, processing of a tissue sample to release individual cells, or isolation

15   from a bodily fluid.

A tissue sample, although similar to a sample of isolated cells, is defined herein as a section of an organ or tissue of the body which typically includes several cell types and/or cytoskeletal structure which holds the cells together. One of skill in the art will appreciate that the term "tissue sample" may be used, in some instances, interchangeably

20   with a "cell sample", although it is preferably used to designate a more complex structure than a cell sample. A tissue sample can be obtained by a biopsy, for example, including by cutting, slicing, or a punch. A bodily fluid sample, like the tissue sample, contains the cells to be evaluated for marker expression or biological activity and/or may contain a soluble biomarker that is secreted by cells, and is a fluid obtained by any method suitable

25   for the particular bodily fluid to be sampled. Bodily fluids suitable for sampling include, but are not limited to, blood, mucous, seminal fluid, saliva, breast milk, bile and urine.

In general, the sample type (*i.e.*, cell, tissue or bodily fluid) is selected based on the accessibility and structure of the organ or tissue to be evaluated for tumor cell growth and/or on what type of cancer is to be evaluated. For example, if the organ/tissue to be

30   evaluated is the breast, the sample can be a sample of epithelial cells from a biopsy (*i.e.*, a cell sample) or a breast tissue sample from a biopsy (a tissue sample). The sample that is most useful in the present invention will be cells, tissues or bodily fluids isolated from a patient by a biopsy or surgery or routine laboratory fluid collection.

Once a sample is obtained from the patient, the sample is evaluated for the detection of the expression of the gene or genes that have been correlated with sensitivity or resistance to an EGFR inhibitor (e.g., gefitinib) of the present invention. For example, as discussed above, any one or more of the genes in Table 1 comprising or expressing a

5   transcript comprising one of SEQ ID NOs:1-194 are useful for detection in the present method.

In one aspect, it may be desirable to select those genes for detection that are particularly highly regulated in gefitinib-sensitive cells versus gefitinib-resistant cells in that they display the largest increases or decreases in expression levels. The detection of

10   such genes can be advantageous because the endpoint may be more clear and require less quantitation. The relative expression levels of the genes identified in the present invention are listed in Table 1, and the genes are ranked in the Table. Therefore, one can easily select subsets of particularly highly regulated genes, or subsets of genes based on some other desired characteristic to provide a more robust, sensitive, or selective assay.

15   In one embodiment, one of skill in the art might choose to detect genes that exhibited a fold increase above background of at least 2. In another embodiment, one of skill in the art might choose to detect genes that exhibited a fold increase or decrease above background of at least 3, and in another embodiment at least 4, and in another embodiment at least 5, and in another embodiment at least 6, and in another embodiment

20   at least 7, and in another embodiment at least 8, and in another embodiment at least 9, and in another embodiment at least 10 or higher fold changes. It is noted that fold increases or decreases are not typically compared from one gene to another, but with reference to the background level for that particular gene.

In one aspect of the method of the present invention, the step of detecting can

25   include the detection of expression of one or more of the genes of this invention. Expression of transcripts and/or proteins is measured by any of a variety of known methods in the art. For RNA expression, methods include but are not limited to: extraction of cellular mRNA and Northern blotting using labeled probes that hybridize to transcripts encoding all or part of one or more of the genes of this invention;

30   amplification of mRNA expressed from one or more of the genes of this invention using gene-specific primers, polymerase chain reaction (PCR), and reverse transcriptase-polymerase chain reaction (RT-PCR), followed by quantitative detection of the product by any of a variety of means; extraction of total RNA from the cells, which is then labeled

and used to probe cDNAs or oligonucleotides encoding all or part of the genes of this invention, arrayed on any of a variety of surfaces; *in situ* hybridization; and detection of a reporter gene.

Methods to measure protein expression levels generally include, but are not limited to: Western blot, immunoblot, enzyme-linked immunosorbant assay (ELISA), radioimmunoassay (RIA), immunoprecipitation, surface plasmon resonance, chemiluminescence, fluorescent polarization, phosphorescence, immunohistochemical analysis, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, microcytometry, microarray, microscopy, fluorescence activated cell sorting (FACS), and flow cytometry, as well as assays based on a property of the protein including but not limited to enzymatic activity or interaction with other protein partners. Binding assays are also well known in the art. For example, a BIAcore machine can be used to determine the binding constant of a complex between two proteins. The dissociation constant for the complex can be determined by monitoring changes in the refractive index with respect to time as buffer is passed over the chip (O'Shannessy et al. *Anal. Biochem.* 212:457 (1993); Schuster et al., *Nature* 365:343 (1993)). Other suitable assays for measuring the binding of one protein to another include, for example, immunoassays such as enzyme linked immunoabsorbent assays (ELISA) and radioimmunoassays (RIA); or determination of binding by monitoring the change in the spectroscopic or optical properties of the proteins through fluorescence, UV absorption, circular dichroism, or nuclear magnetic resonance (NMR).

Nucleic acid arrays are particularly useful for detecting the expression of the genes of the present invention. The production and application of high-density arrays in gene expression monitoring have been disclosed previously in, for example, WO 97/10365; WO 92/10588; U.S. Patent No. 6,040,138; U.S. 5,445,934; or WO95/35505, all of which are incorporated herein by reference in their entireties. Also for examples of arrays, see Hacia *et al.* (1996) *Nature Genetics* 14:441-447; Lockhart *et al.* (1996) Nature Biotechnol. 14:1675-1680; and De Risi *et al.* (1996) *Nature Genetics* 14:457-460. In general, in an array, an oligonucleotide, a cDNA, or genomic DNA, that is a portion of a known gene occupies a known location on a substrate. A nucleic acid target sample is hybridized with an array of such oligonucleotides and then the amount of target nucleic acids hybridized to each probe in the array is quantified. One preferred quantifying method is to use confocal microscope and fluorescent labels. The *Affymetrix GeneChip*™

Array system (Affymetrix, Santa Clara, Calif.) and the *Atlas*^{TM} *Human cDNA Expression Array* system are particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar systems or other effectively equivalent detection methods can also be used. In a particularly preferred embodiment, one can use the knowledge of the genes described herein to design novel arrays of polynucleotides, cDNAs or genomic DNAs for screening methods described herein. Such novel pluralities of polynucleotides are contemplated to be a part of the present invention and are described in detail below.

Suitable nucleic acid samples for screening on an array contain transcripts of interest or nucleic acids derived from the transcripts of interest. As used herein, a nucleic acid derived from a transcript refers to a nucleic acid for whose synthesis the mRNA transcript or a subsequence thereof has ultimately served as a template. Thus, a cDNA reverse transcribed from a transcript, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from the transcript and detection of such derived products is indicative of the presence and/or abundance of the original transcript in a sample. Thus, suitable samples include, but are not limited to, transcripts of the gene or genes, cDNA reverse transcribed from the transcript, cRNA transcribed from the cDNA, DNA amplified from the genes, RNA transcribed from amplified DNA, and the like. Preferably, the nucleic acids for screening are obtained from a homogenate of cells or tissues or other biological samples. Preferably, such sample is a total RNA preparation of a biological sample. More preferably in some embodiments, such a nucleic acid sample is the total mRNA isolated from a biological sample. Biological samples may be of any biological tissue or fluid or cells from any organism. Frequently the sample will be a "clinical sample" which is a sample derived from a patient, such as a lung tumor sample from a patient. Typical clinical samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. Biological samples may also include sections of tissues, such as frozen sections or formalin fixed sections taken for histological purposes.

In one embodiment, it is desirable to amplify the nucleic acid sample prior to hybridization. One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids to achieve

quantitative amplification. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. The high-density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid. Other suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), transcription amplification (Kwoh, et al., Proc. Natl. Acad. Sci. USA, 86: 1173 (1989)), and self-sustained sequence replication (Guatelli, et al, Proc. Nat. Acad. Sci. USA, 87: 1874 (1990)).

Nucleic acid hybridization simply involves contacting a probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing. As used herein, hybridization conditions refer to standard hybridization conditions under which nucleic acid molecules are used to identify similar nucleic acid molecules. Such standard conditions are disclosed, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety (see specifically, pages 9.31-9.62). In addition, formulae to calculate the appropriate hybridization and wash conditions to achieve hybridization permitting varying degrees of mismatch of nucleotides are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284; Meinkoth et al., *ibid.*, is incorporated by reference herein in its entirety. Nucleic acids that do not form hybrid duplexes are washed away from the hybridized nucleic acids and the hybridized nucleic acids can then be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

High stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 90% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 10% or less mismatch of nucleotides). One of skill in the art can use the formulae in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284 (incorporated herein by reference in its entirety) to calculate the appropriate hybridization and wash conditions to achieve these particular levels of nucleotide mismatch. Such conditions will vary, depending on whether DNA:RNA or DNA:DNA hybrids are being formed. Calculated melting temperatures for DNA:DNA hybrids are 10°C less than for DNA:RNA hybrids. In particular embodiments, stringent hybridization conditions for DNA:DNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na$^+$) at a temperature of between about 20°C and about 35°C, more preferably, between about 28°C and about 40°C, and even more preferably, between about 35°C and about 45°C. In particular embodiments, stringent hybridization conditions for DNA:RNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na$^+$) at a temperature of between about 30°C and about 45°C, more preferably, between about 38°C and about 50°C, and even more preferably, between about 45°C and about 55°C. These values are based on calculations of a melting temperature for molecules larger than about 100 nucleotides, 0% formamide and a G + C content of about 40%. Alternatively, T_m can be calculated empirically as set forth in Sambrook et al., *supra*, pages 9.31 to 9.62.

The hybridized nucleic acids are detected by detecting one or more labels attached to the sample nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill in the art. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads.TM.), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected

using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

The term "quantifying" or "quantitating" when used in the context of quantifying transcription levels of a gene can refer to absolute or to relative quantification. Absolute quantification may be accomplished by inclusion of known concentration(s) of one or more target nucleic acids and referencing the hybridization intensity of unknowns with the known target nucleic acids (e.g. through generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication, transcription level.

In one aspect of the present method, *in vitro* cell based assays may be designed to screen for compounds that affect the regulation of genes at either the transcriptional or translational level. One, two or more promoters of the genes of this invention can be used to screen unknown compounds for activity on a given target. Promoters of the selected genes can be linked to any of several reporters (including but not limited to chloramphenicol acetyl transferase, or luciferase) that measure transcriptional read-out. The promoters can be tested as pure DNA, or as DNA bound to chromatin proteins.

In one aspect of the present method, the step of detecting can include detecting the expression of one or more genes of the invention in intact animals or tissues obtained from such animals. Mammalian (i.e. mouse, rat, monkey) or non-mammalian (i.e. chicken) species can be the test animals. Sample tissues from a patient can also be screened. The tissues to be surveyed can be either normal or malignant tissues. The presence and quantity of endogenous mRNA or protein expression of one or more of the genes of this invention can be measured in those tissues. The gene markers can be measured in tissues that are fresh, frozen, fixed or otherwise preserved. They can be measured in cytoplasmic or nuclear organ-, tissue- or cell-extracts; or in cell membranes including but not limited to plasma, cytoplasmic, mitochondrial, golgi or nuclear membranes; in the nuclear matrix; or in cellular organelles and their extracts including but not limited to ribosomes, nuclei, nucleoli, mitochondria, or golgi. Assays for endogenous expression of mRNAs or proteins encoded by the genes of this invention can

be performed as described above. Alternatively, intact transgenic animals can be generated for screening for research or validation purposes.

Preferably, a gene identified as being upregulated or downregulated in a test cell according to the invention (including a sample tumor cell to be screened) is regulated in the same direction and to at least about 5%, and more preferably at least about 10%, and more preferably at least 20%, and more preferably at least 25%, and more preferably at least 30%, and more preferably at least 35%, and more preferably at least 40%, and more preferably at least 45%, and more preferably at least 50%, and preferably at least 55%, and more preferably at least 60%, and more preferably at least 65%, and more preferably at least 70%, and more preferably at least 75%, and more preferably at least 80%, and more preferably at least 85%, and more preferably at least 90%, and more preferably at least 95%, and more preferably of 100%, or any percentage change between 5% and higher in 1% increments (*i.e.*, 5%, 6%, 7%, 8%...), of the level of expression of the gene that is seen in established or confirmed gefitinib-sensitive or gefitinib-resistant cells. A gene identified as being upregulated or downregulated in a test cell according to the invention can also be regulated in the same direction and to a higher level than the level of expression of the gene that is seen in established or confirmed gefitinib-sensitive or gefitinib-resistant cells.

The values obtained from the test and/or control samples are statistically processed using any suitable method of statistical analysis to establish a suitable baseline level using methods standard in the art for establishing such values. Statistical significance according to the present invention should be at least $p<0.05$.

It will be appreciated by those of skill in the art that differences between the expression of genes in sensitive versus resistant cells may be small or large. Some small differences may be very reproducible and therefore nonetheless useful. For other purposes, large differences may be desirable for ease of detection of the activity. It will be therefore appreciated that the exact boundary between what is called a positive result and a negative result can shift, depending on the goal of the screening assay and the genes to be screened. For some assays it may be useful to set threshold levels of change. One of skill in the art can readily determine the criteria for screening of cells given the information provided herein.

The presence and quantity of each gene marker can be measured in primary tumors, metastatic tumors, locally recurring tumors, ductal carcinomas in situ, or other

tumors. The markers can be measured in solid tumors that are fresh, frozen, fixed or otherwise preserved. They can be measured in cytoplasmic or nuclear tumor extracts; or in tumor membranes including but not limited to plasma, mitochondrial, golgi or nuclear membranes; in the nuclear matrix; or in tumor cell organelles and their extracts including but not limited to ribosomes, nuclei, mitochondria, golgi.

The level of expression of the gene or genes detected in the test or patient sample f the invention is compared to a baseline or control level of expression of that gene. More specifically, according to the present invention, a "baseline level" is a control level of biomarker expression against which a test level of biomarker expression (*i.e.*, in the test sample) can be compared. In the present invention, the control level of biomarker expression can be the expression level of the gene or genes in a control cell that is sensitive to the EGFR inhibitor, and/or the expression level of the gene or genes in a control cell that is resistant to the EGFR inhibitor. Other controls may also be included in the assay. In one embodiment, the control is established in an autologous control sample obtained from the patient. The autologous control sample can be a sample of isolated cells, a tissue sample or a bodily fluid sample, and is preferably a cell sample or tissue sample. According to the present invention, and as used in the art, the term "autologous" means that the sample is obtained from the same patient from which the sample to be evaluated is obtained. The control sample should be of or from the same cell type and preferably, the control sample is obtained from the same organ, tissue or bodily fluid as the sample to be evaluated, such that the control sample serves as the best possible baseline for the sample to be evaluated. In one embodiment, control expression levels of the gene or genes that has been correlated with sensitivity and/or resistance to the EGFR inhibitor has been predetermined, such as in Table 1. Such a form of stored information can include, for example, but is not limited to, a reference chart, listing or electronic file of gene expression levels and profiles for EGFR inhibitor sensitive and/or EGFR inhibitor resistant biomarker expression, or any other source of data regarding baseline biomarker expression that is useful in the method of the invention. Therefore, it can be determined, based on the control or baseline level of biomarker expression or biological activity, whether the expression level of a gene or genes in a patient sample is/are more statistically significantly similar to the baseline for EGFR resistance or EGFR sensitivity.

A profile of individual gene markers, including a matrix of two or more markers, can be generated by one or more of the methods described above. According to the

present invention, a profile of the genes in a tissue sample refers to a reporting of the expression level of a given gene from Table 1, and includes a classification of the gene with regard to how the gene is regulated in gefitinib-sensitive versus gefitinib-resistant cells. The data can be reported as raw data, and/or statistically analyzed by any of a

5   variety of methods, and/or combined with any other prognostic marker(s).

Another embodiment of the present invention relates to a plurality of polynucleotides for the detection of the expression of genes as described herein. The plurality of polynucleotides consists of polynucleotides that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes listed in Table 1 or otherwise

10   identified as being useful according to the present invention (e.g., other genes correlated with sensitivity or resistance to gefitinib or another EGFR inhibitor), and is therefore distinguished from previously known nucleic acid arrays and primer sets. The plurality of polynucleotides within the above-limitation includes at least two or more polynucleotides that are complementary to RNA transcripts, or nucleotides derived therefrom, of one or

15   more genes identified by the present inventors and listed in Table 1. Preferably, the plurality of polynucleotides is capable of detecting expression of at least two, and more preferably at least five, and more preferably at least 10, and more preferably at least 25, and more preferably at least 50, and more preferably at least 100, and more preferably at least 150, and more preferably all of the genes (or any number in between two and all of

20   the genes, in whole increments) in a panel of genes correlated with EGFR inhibitor sensitivity and/or resistance, such as all of the genes listed in Table 1.

In one embodiment, it is contemplated that additional genes that are not regulated differently in gefitinib-sensitive versus gefitinib-resistant cells can be added to the plurality of polynucleotides. Such genes would not be random genes, or large groups of

25   unselected human genes, as are commercially available now, but rather, would be specifically selected to complement the sets of genes identified by the present invention. For example, one of skill in the art may wish to add to the above-described plurality of genes one or more genes that are of relevance because they are expressed by a particular tissue of interest (e.g., lung tissue), are associated with a particular disease or condition of

30   interest (e.g., NSCLC), or are associated with a particular cell, tissue or body function (e.g., angiogenesis). The development of additional pluralities of polynucleotides (and antibodies, as disclosed below), which include both the above-described plurality and

such additional selected polynucleotides, are explicitly contemplated by the present invention.

According to the present invention, a plurality of polynucleotides refers to at least 2, and more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in increments of one, up to any suitable number of polynucleotides, including at least 100, 500, 1000, 10^4, 10^5, or at least 10^6 or more polynucleotides.

In one embodiment, the polynucleotide probes are conjugated to detectable markers. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads.TM.), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^3H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Preferably, the polynucleotide probes are immobilized on a substrate.

In one embodiment, the polynucleotide probes are hybridizable array elements in a microarray or high density array. Nucleic acid arrays are well known in the art and are described for use in comparing expression levels of particular genes of interest, for example, in U.S. Patent No. 6,177,248, which is incorporated herein by reference in its entirety. Nucleic acid arrays are suitable for quantifying a small variations in expression levels of a gene in the presence of a large population of heterogeneous nucleic acids. Knowing the identity of the genes of the present invention, nucleic acid arrays can be fabricated either by *de novo* synthesis on a substrate or by spotting or transporting nucleic acid sequences onto specific locations of substrate. Nucleic acids are purified and/or isolated from biological materials, such as a bacterial plasmid containing a cloned segment of sequence of interest. It is noted that all of the genes identified by the present invention have been previously sequenced, at least in part, such that oligonucleotides suitable for the identification of such nucleic acids can be produced. The database accession number for each of the genes identified by the present inventors is provided in

Table 1. Suitable nucleic acids are also produced by amplification of template, such as by polymerase chain reaction or in vitro transcription.

Synthesized oligonucleotide arrays are particularly preferred for this aspect of the invention. Oligonucleotide arrays have numerous advantages, as opposed to other methods, such as efficiency of production, reduced intra- and inter array variability, increased information content and high signal-to-noise ratio.

One of skill in the art will appreciate that an enormous number of array designs are suitable for the practice of this invention. An array will typically include a number of probes that specifically hybridize to the sequences of interest. In addition, in a preferred embodiment, the array will include one or more control probes. The high-density array chip includes "test probes." Test probes could be oligonucleotides that range from about 5 to about 45 or 5 to about 500 nucleotides (including any whole number increment in between), more preferably from about 10 to about 40 nucleotides and most preferably from about 15 to about 40 nucleotides in length. In other particularly preferred embodiments the probes are 20 or 25 nucleotides in length. In another preferred embodiments, test probes are double or single strand DNA sequences. DNA sequences are isolated or cloned from natural sources or amplified from natural sources using natural nucleic acids as templates, or produced synthetically. These probes have sequences complementary to particular subsequences of the genes whose expression they are designed to detect. Thus, the test probes are capable of specifically hybridizing to the target nucleic acid they are to detect.

Another embodiment of the present invention relates to a plurality of antibodies, or antigen binding fragments thereof, for the detection of the expression of genes according to the present invention. The plurality of antibodies, or antigen binding fragments thereof, consists of antibodies, or antigen binding fragments thereof, that selectively bind to proteins encoded by genes described herein. According to the present invention, a plurality of antibodies, or antigen binding fragments thereof, refers to at least 2, and more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in increments of one, up to any suitable number of antibodies, or antigen binding fragments thereof, including at least 100, 500, or at least 1000 antibodies, or antigen binding fragments thereof.

The invention also extends to non-antibody polypeptides, sometimes referred to as binding partners or antigen binding peptides, that have been designed to bind specifically to, and either activate or inhibit as appropriate, a target protein. Examples of the design of such polypeptides, which possess a prescribed ligand specificity are given in Beste et

5   al. (*Proc. Natl. Acad. Sci.* 96:1898-1903, 1999), incorporated herein by reference in its entirety.

Limited digestion of an immunoglobulin with a protease may produce two fragments. An antigen binding fragment is referred to as an Fab, an Fab', or an F(ab')$_2$ fragment. A fragment lacking the ability to bind to antigen is referred to as an Fc

10   fragment. An Fab fragment comprises one arm of an immunoglobulin molecule containing a L chain ($V_L + C_L$ domains) paired with the V_H region and a portion of the C_H region (CH1 domain). An Fab' fragment corresponds to an Fab fragment with part of the hinge region attached to the CH1 domain. An F(ab')$_2$ fragment corresponds to two Fab' fragments that are normally covalently linked to each other through a di-sulfide bond,

15   typically in the hinge regions.

Isolated antibodies of the present invention can include serum containing such antibodies, or antibodies that have been purified to varying degrees. Whole antibodies of the present invention can be polyclonal or monoclonal. Alternatively, functional equivalents of whole antibodies, such as antigen binding fragments in which one or more

20   antibody domains are truncated or absent (e.g., Fv, Fab, Fab', or F(ab)$_2$ fragments), as well as genetically-engineered antibodies or antigen binding fragments thereof, including single chain antibodies or antibodies that can bind to more than one epitope (e.g., bi-specific antibodies), or antibodies that can bind to one or more different antigens (e.g., bi- or multi-specific antibodies), may also be employed in the invention.

25   Generally, in the production of an antibody, a suitable experimental animal, such as, for example, but not limited to, a rabbit, a sheep, a hamster, a guinea pig, a mouse, a rat, or a chicken, is exposed to an antigen against which an antibody is desired. Typically, an animal is immunized with an effective amount of antigen that is injected into the animal. An effective amount of antigen refers to an amount needed to induce

30   antibody production by the animal. The animal's immune system is then allowed to respond over a pre-determined period of time. The immunization process can be repeated until the immune system is found to be producing antibodies to the antigen. In order to obtain polyclonal antibodies specific for the antigen, serum is collected from the animal

that contains the desired antibodies (or in the case of a chicken, antibody can be collected from the eggs). Such serum is useful as a reagent. Polyclonal antibodies can be further purified from the serum (or eggs) by, for example, treating the serum with ammonium sulfate.

5   Monoclonal antibodies may be produced according to the methodology of Kohler and Milstein (*Nature* 256:495-497, 1975). For example, B lymphocytes are recovered from the spleen (or any suitable tissue) of an immunized animal and then fused with myeloma cells to obtain a population of hybridoma cells capable of continual growth in suitable culture medium. Hybridomas producing the desired antibody are selected by

10   testing the ability of the antibody produced by the hybridoma to bind to the desired antigen.

Finally, any of the genes of this invention, or their RNA or protein products, can serve as targets for therapeutic strategies. For example, neutralizing antibodies could be directed against one of the protein products of a selected gene, expressed on the surface of

15   a tumor cell. Alternatively, regulatory compounds that regulate (e.g., upregulate or downregulate) the expression and/or biological activity of a target gene (whether the product is intracellular, membrane or secreted), can be identified and/or designed using the genes described herein. For example, in one aspect, a method of using the genes described herein as a target includes the steps of: (a) contacting a test compound with a

20   cell that expresses at least one gene, wherein said gene is selected from any one of the genes comprising, or expressing a transcript comprising, a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1-194; and (b) identifying compounds, wherein the compounds can include: (i) compounds that increase the expression or activity of the gene or genes in (a), or the proteins encoded thereby, that are correlated

25   with sensitivity to gefitinib; and (ii) compounds that decrease the expression or activity of genes in (a), or the proteins encoded thereby, that are correlated with resistance to gefitinib. The compounds are thereby identified as having the potential to enhance the efficacy of EGFR inhibitors.

The period of contact with the compound being tested can be varied depending on

30   the result being measured, and can be determined by one of skill in the art. As used herein, the term "contact period" refers to the time period during which cells are in contact with the compound being tested. The term "incubation period" refers to the entire time during which cells are allowed to grow prior to evaluation, and can be inclusive of

the contact period. Thus, the incubation period includes all of the contact period and may include a further time period during which the compound being tested is not present but during which expression of genes is allowed to continue prior to scoring. Methods to evaluate gene expression in a cell according to the present invention have been described

5   previously herein.

If a suitable therapeutic compound is identified using the methods and genes of the present invention, a composition can be formulated. A composition, and particularly a therapeutic composition, of the present invention generally includes the therapeutic compound and a carrier, and preferably, a pharmaceutically acceptable carrier.

10   According to the present invention, a "pharmaceutically acceptable carrier" includes pharmaceutically acceptable excipients and/or pharmaceutically acceptable delivery vehicles, which are suitable for use in administration of the composition to a suitable *in vitro, ex vivo* or *in vivo* site. A suitable *in vitro, in vivo* or *ex vivo* site is preferably a tumor cell. In some embodiments, a suitable site for delivery is a site of inflammation,

15   near the site of a tumor, or a site of any other disease or condition in which regulation of the genes identified herein can be beneficial. Preferred pharmaceutically acceptable carriers are capable of maintaining a compound, a protein, a peptide, nucleic acid molecule or mimetic (drug) according to the present invention in a form that, upon arrival of the compound, protein, peptide, nucleic acid molecule or mimetic at the cell target in a

20   culture or in patient, the compound, protein, peptide, nucleic acid molecule or mimetic is capable of interacting with its target.

Suitable excipients of the present invention include excipients or formularies that transport or help transport, but do not specifically target a composition to a cell (also referred to herein as non-targeting carriers). Examples of pharmaceutically acceptable

25   excipients include, but are not limited to water, phosphate buffered saline, Ringer's solution, dextrose solution, serum-containing solutions, Hank's solution, other aqueous physiologically balanced solutions, oils, esters and glycols. Aqueous carriers can contain suitable auxiliary substances required to approximate the physiological conditions of the recipient, for example, by enhancing chemical stability and isotonicity.

30   Suitable auxiliary substances include, for example, sodium acetate, sodium chloride, sodium lactate, potassium chloride, calcium chloride, and other substances used to produce phosphate buffer, Tris buffer, and bicarbonate buffer. Auxiliary substances can also include preservatives, such as thimerosal, m- or o-cresol, formalin and benzol

33

alcohol. Compositions of the present invention can be sterilized by conventional methods and/or lyophilized.

One type of pharmaceutically acceptable carrier includes a controlled release formulation that is capable of slowly releasing a composition of the present invention into a patient or culture. As used herein, a controlled release formulation comprises a compound of the present invention (e.g., a protein (including homologues), a drug, an antibody, a nucleic acid molecule, or a mimetic) in a controlled release vehicle. Suitable controlled release vehicles include, but are not limited to, biocompatible polymers, other polymeric matrices, capsules, microcapsules, microparticles, bolus preparations, osmotic pumps, diffusion devices, liposomes, lipospheres, and transdermal delivery systems. Other carriers of the present invention include liquids that, upon administration to a patient, form a solid or a gel *in situ*. Preferred carriers are also biodegradable (i.e., bioerodible). When the compound is a recombinant nucleic acid molecule, suitable delivery vehicles include, but are not limited to liposomes, viral vectors or other delivery vehicles, including ribozymes. Natural lipid-containing delivery vehicles include cells and cellular membranes. Artificial lipid-containing delivery vehicles include liposomes and micelles. A delivery vehicle of the present invention can be modified to target to a particular site in a patient, thereby targeting and making use of a compound of the present invention at that site. Suitable modifications include manipulating the chemical formula of the lipid portion of the delivery vehicle and/or introducing into the vehicle a targeting agent capable of specifically targeting a delivery vehicle to a preferred site, for example, a preferred cell type. Other suitable delivery vehicles include gold particles, poly-L-lysine/DNA-molecular conjugates, and artificial chromosomes.

A pharmaceutically acceptable carrier which is capable of targeting is herein referred to as a "delivery vehicle." Delivery vehicles of the present invention are capable of delivering a composition of the present invention to a target site in a patient. A "target site" refers to a site in a patient to which one desires to deliver a composition. For example, a target site can be any cell which is targeted by direct injection or delivery using liposomes, viral vectors or other delivery vehicles, including ribozymes and antibodies. Examples of delivery vehicles include, but are not limited to, artificial and natural lipid-containing delivery vehicles, viral vectors, and ribozymes. Natural lipid-containing delivery vehicles include cells and cellular membranes. Artificial lipid-containing delivery vehicles include liposomes and micelles. A delivery vehicle of the

present invention can be modified to target to a particular site in a subject, thereby targeting and making use of a compound of the present invention at that site. Suitable modifications include manipulating the chemical formula of the lipid portion of the delivery vehicle and/or introducing into the vehicle a compound capable of specifically targeting a delivery vehicle to a preferred site, for example, a preferred cell type. Specifically, targeting refers to causing a delivery vehicle to bind to a particular cell by the interaction of the compound in the vehicle to a molecule on the surface of the cell. Suitable targeting compounds include ligands capable of selectively (i.e., specifically) binding another molecule at a particular site. Examples of such ligands include antibodies, antigens, receptors and receptor ligands. Manipulating the chemical formula of the lipid portion of the delivery vehicle can modulate the extracellular or intracellular targeting of the delivery vehicle. For example, a chemical can be added to the lipid formula of a liposome that alters the charge of the lipid bilayer of the liposome so that the liposome fuses with particular cells having particular charge characteristics.

Another preferred delivery vehicle comprises a viral vector. A viral vector includes an isolated nucleic acid molecule useful in the present invention, in which the nucleic acid molecules are packaged in a viral coat that allows entrance of DNA into a cell. A number of viral vectors can be used, including, but not limited to, those based on alphaviruses, poxviruses, adenoviruses, herpesviruses, lentiviruses, adeno-associated viruses and retroviruses.

A composition can be delivered to a cell culture or patient by any suitable method. Selection of such a method will vary with the type of compound being administered or delivered (i.e., compound, protein, peptide, nucleic acid molecule, or mimetic), the mode of delivery (i.e., *in vitro*, *in vivo*, *ex vivo*) and the goal to be achieved by administration/delivery of the compound or composition. According to the present invention, an effective administration protocol (i.e., administering a composition in an effective manner) comprises suitable dose parameters and modes of administration that result in delivery of a composition to a desired site (i.e., to a desired cell) and/or in the desired regulatory event.

Administration routes include *in vivo*, *in vitro* and *ex vivo* routes. *In vivo* routes include, but are not limited to, oral, nasal, intratracheal injection, inhaled, transdermal, rectal, and parenteral routes. Preferred parenteral routes can include, but are not limited to, subcutaneous, intradermal, intravenous, intramuscular and intraperitoneal routes.

Intravenous, intraperitoneal, intradermal, subcutaneous and intramuscular administrations can be performed using methods standard in the art. Aerosol (inhalation) delivery can also be performed using methods standard in the art (see, for example, Stribling et al., *Proc. Natl. Acad. Sci. USA* 189:11277-11281, 1992, which is incorporated herein by reference in its entirety). Oral delivery can be performed by complexing a therapeutic composition of the present invention to a carrier capable of withstanding degradation by digestive enzymes in the gut of an animal. Examples of such carriers, include plastic capsules or tablets, such as those known in the art. Direct injection techniques are particularly useful for suppressing graft rejection by, for example, injecting the composition into the transplanted tissue, or for site-specific administration of a compound, such as at the site of a tumor. *Ex vivo* refers to performing part of the regulatory step outside of the patient, such as by transfecting a population of cells removed from a patient with a recombinant molecule comprising a nucleic acid sequence encoding a protein according to the present invention under conditions such that the recombinant molecule is subsequently expressed by the transfected cell, and returning the transfected cells to the patient. *In vitro* and *ex vivo* routes of administration of a composition to a culture of host cells can be accomplished by a method including, but not limited to, transfection, transformation, electroporation, microinjection, lipofection, adsorption, protoplast fusion, use of protein carrying agents, use of ion carrying agents, use of detergents for cell permeabilization, and simply mixing (e.g., combining) a compound in culture with a target cell.

In the method of the present invention, a therapeutic compound, as well as compositions comprising such compounds, can be administered to any organism, and particularly, to any member of the Vertebrate class, Mammalia, including, without limitation, primates, rodents, livestock and domestic pets. Livestock include mammals to be consumed or that produce useful products (e.g., sheep for wool production). Preferred mammals to protect include humans. Typically, it is desirable to obtain a therapeutic benefit in a patient. A therapeutic benefit is not necessarily a cure for a particular disease or condition, but rather, preferably encompasses a result which can include alleviation of the disease or condition, elimination of the disease or condition, reduction of a symptom associated with the disease or condition, prevention or alleviation of a secondary disease or condition resulting from the occurrence of a primary disease or condition, and/or prevention of the disease or condition. As used herein, the phrase "protected from a

disease" refers to reducing the symptoms of the disease; reducing the occurrence of the disease, and/or reducing the severity of the disease. Protecting a patient can refer to the ability of a composition of the present invention, when administered to a patient, to prevent a disease from occurring and/or to cure or to alleviate disease symptoms, signs or

5   causes. As such, to protect a patient from a disease includes both preventing disease occurrence (prophylactic treatment) and treating a patient that has a disease (therapeutic treatment) to reduce the symptoms of the disease. A beneficial effect can easily be assessed by one of ordinary skill in the art and/or by a trained clinician who is treating the patient. The term, "disease" refers to any deviation from the normal health of a mammal

10   and includes a state when disease symptoms are present, as well as conditions in which a deviation (e.g., infection, gene mutation, genetic defect, etc.) has occurred, but symptoms are not yet manifested.

Various aspects of the invention are described in the following examples;

15   however, the following examples are provided for the purpose of illustration and are not intended to limit the scope of the present invention.

Examples

Example 1

       The following example describes the identification of a biomarker panel that

20   discriminates gefitinib-sensitive cell lines from gefitinib-resistant cell lines.

       Methods: Gefitinib sensitivity was determined in 18 NSCLC cell lines using MTT assays. Cell lines were classified as gefitinib sensitive ($IC_{50}<1\mu M$), resistant ($IC_{50}>10\mu M$) or intermediate sensitivity ($10\mu M<IC_{50}>1$). Oligonucleotide gene arrays (Affymetrix® Human Genome U133 set, 39,000 genes) were done on 10 cell lines. Three

25   distinct filtration and normalization algorithms to process the expression data were used, and a list of genes were generated that were both statistically significant (unadjusted p=0.001 cutoff) and corrected for false positive occurrence. This approach was used in combination with 5 distinct machine learning algorithms used to build a test set for predictor genes that were successful for 100% of the test cases. The best discriminators

30   (>3 fold difference in expression between sensitive and resistant cell lines) were selected for Real-time RT-PCR.

       Results: A list of genes was generated initially from the Affymetrix array analysis. By using the mathematical algorithm, 14 different candidate genes were

selected for RT-PCR. Twelve of the 14 genes were verified to discriminate between sensitive and resistant cell lines by Real-time RT-PCR.

Conclusion: Based on NSCLC cell line studies it was possible to identify genes which strongly discriminated gefitinib (Iressa) sensitive cell lines from the resistant ones. The genes are ranked in Table 1. This entire biomarker panel is of significant value for selecting NSCLC patients for gefitinib treatment.

Table 1.

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
202286_s_at	0.00000005	3.8	9893.5	TACSTD2	SEQ ID NO:12	tumor-associated calcium signal transducer 2
202489_s_at	0.00000005	25.8	2372.6	FXYD3	SEQ ID NO:16	FXYD domain containing ion transport regulator 3
213285_at	0.00000005	8.0	1739.3	TMEM30B	SEQ ID NO:73	transmembrane protein 30B
218186_at	0.00000005	3.6	2295.0	RAB25	SEQ ID NO:83	RAB25, member RAS oncogene family
235515_at	0.00000005	6.4	287.6	FLJ36445	SEQ ID NO:168	hypothetical protein FLJ36445
235988_at	0.00000005	11.3	345.7	GPR110	SEQ ID NO:170	G protein-coupled receptor 110
238689_at	0.00000005	5.4	2210.5	GPR110	SEQ ID NO:177	G protein-coupled receptor 110
232165_at	0.00000010	4.6	244.0	EPPK1	SEQ ID NO:164	epiplakin 1
240633_at	0.00000010	6.2	61.2	FLJ33718	SEQ ID NO:182	hypothetical protein FLJ33718
219525_at	0.00000020	179.3	6.1	FLJ10847	SEQ ID NO:93	hypothetical protein FLJ10847
229599_at	0.00000020	5.9	112.8		SEQ ID NO:154	Clone IMAGE:5166045, Mrna
203397_s_at	0.00000030	10.1	1128.6	GALNT3	SEQ ID NO:28	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 (GalNAc-T3)
232164_s_at	0.00000030	5.8	411.1	EPPK1	SEQ ID NO:163	epiplakin 1
212813_at	0.00000060	163.8	7.9	JAM3	SEQ ID NO:71	junctional adhesion molecule 3
227134_at	0.00000160	14.2	705.7	SYTL1	SEQ ID NO:143	synaptotagmin-like 1
236489_at	0.00000170	8.2	498.5		SEQ ID NO:171	
235651_at	0.00000480	3.9	98.2		SEQ ID NO:169	
238439_at	0.00000700	7.7	537.6	ANKRD22	SEQ ID NO:173	ankyrin repeat domain 22
219388_at	0.00000730	19.3	342.1	TFCP2L3	SEQ ID NO:91	transcription factor CP2-like 3
227985_at	0.00000820	5.0	179.9		SEQ ID NO:146	
227450_at	0.00000890	5.1	509.7	FLJ32115	SEQ ID NO:144	hypothetical protein FLJ32115
203256_at	0.00000980	13.4	2223.0	CDH3	SEQ ID NO:23	cadherin 3, type 1, P-cadherin (placental)
220318_at	0.00000980	4.4	44.7	EPN3	SEQ ID NO:100	epsin 3
202525_at	0.00001030	7.8	1194.6	PRSS8	SEQ ID NO:17	protease, serine, 8 (prostasin)
227803_at	0.00001080	7.8	206.1	ENPP5	SEQ ID NO:145	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative function)
206884_s_at	0.00001200	12.8	822.7	SCEL	SEQ ID NO:49	Sciellin
223895_s_at	0.00001290	13.8	183.6	EPN3	SEQ ID NO:119	epsin 3
238493_at	0.00001650	7.3	18.5	ZNF506	SEQ ID NO:174	zinc finger protein 506
224913_s_at	0.00001960	2703.8	1081.5	TIMM50	SEQ ID NO:122	translocase of inner mitochondrial membrane 50 homolog (yeast)

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
201428_at	0.00002330	90.3	3416.4	CLDN4	SEQ ID NO:5	claudin 4
216641_s_at	0.00003760	26.8	423.5	LAD1	SEQ ID NO:78	ladinin 1
231929_at	0.00003910	31.0	340.7		SEQ ID NO:159	MRNA; cDNA DKFZp586O0724 (from clone DKFZp586O0724)
212764_at	0.00003930	320.0	9.2	TCF8	SEQ ID NO:70	transcription factor 8 (represses interleukin 2 expression)
238778_at	0.00004080	15.0	106.1	MPP7	SEQ ID NO:178	membrane protein, palmitoylated 7 (MAGUK p55 subfamily member 7)
202641_at	0.00004360	2011.3	933.3	ARL3	SEQ ID NO:19	ADP-ribosylation factor-like 3
212233_at	0.00004550	2005.7	137.0	MAP1B	SEQ ID NO:66	microtubule-associated protein 1B
224232_s_at	0.00004560	1054.1	438.3	PX19	SEQ ID NO:120	px19-like protein
226905_at	0.00004590	240.2	14.0	MGC45871	SEQ ID NO:142	hypothetical protein MGC45871
218553_s_at	0.00004620	177.0	38.2	KCTD15	SEQ ID NO:84	potassium channel tetramerisation domain containing 15
215218_s_at	0.00004830	368.6	142.8	C19orf14	SEQ ID NO:77	chromosome 19 open reading frame 14
203287_at	0.00004920	23.4	505.0	LAD1	SEQ ID NO:24	ladinin 1
209114_at	0.00005560	43.7	717.2	TSPAN-1	SEQ ID NO:57	tetraspan 1
230076_at	0.00005660	21.2	120.1		SEQ ID NO:155	
218677_at	0.00005710	21.5	966.3	S100A14	SEQ ID NO:85	S100 calcium binding protein A14
236616_at	0.00005810	17.8	32.9		SEQ ID NO:172	CDNA FLJ41623 fis, clone CTONG3009227
205014_at	0.00006280	13.4	491.2	FGFBP1	SEQ ID NO:40	fibroblast growth factor binding protein 1
200720_s_at	0.00006360	1089.8	391.9	ACTR1A	SEQ ID NO:2	ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)
224326_s_at	0.00006750	499.6	135.5	RNF134	SEQ ID NO:121	ring finger protein 134 /// ring finger protein 134
242138_at	0.00006800	207.4	6.9	DLX1	SEQ ID NO:184	distal-less homeo box 1
90265_at	0.00007110	145.0	1117.7	CENTA1	SEQ ID NO:193	centaurin, alpha 1
222360_at	0.00007190	97.8	21.2	CGI-30	SEQ ID NO:108	CGI-30 protein
208393_s_at	0.00007530	1370.0	596.5	RAD50	SEQ ID NO:53	RAD50 homolog (S. cerevisiae)
226403_at	0.00007930	22.5	680.1	TMC4	SEQ ID NO:136	transmembrane channel-like 4
232056_at	0.00008450	9.8	141.7	SCEL	SEQ ID NO:160	Sciellin
207655_s_at	0.00008700	7.1	71.1	BLNK	SEQ ID NO:51	B-cell linker
228683_s_at	0.00009450	101.5	18.5	KCTD15	SEQ ID NO:148	potassium channel tetramerisation domain containing 15

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
204160_s_at	0.00009570	23.9	314.8	ENPP4	SEQ ID NO:36	Ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)
202454_s_at	0.00009860	16.3	1266.2	ERBB3	SEQ ID NO:15	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
232151_at	0.00010020	8.5	295.7		SEQ ID NO:162	MRNA full length insert cDNA clone EUROIMAGE 2344436
205073_at	0.00010350	30.8	136.8	CYP2J2	SEQ ID NO:41	cytochrome P450, family 2, subfamily J, polypeptide 2
225658_at	0.00011660	167.1	516.3	LOC339745	SEQ ID NO:127	hypothetical protein LOC339745
219150_s_at	0.00012240	30.9	200.1	CENTA1	SEQ ID NO:90	centaurin, alpha 1
228882_at	0.00012370	152.7	10.4	TUB	SEQ ID NO:150	tubby homolog (mouse)
222857_s_at	0.00012430	17.2	344.7	KCNMB4	SEQ ID NO:113	potassium large conductance calcium-activated channel, subfamily M, beta member 4
55662_at	0.00013490	84.7	31.7	C10orf76	SEQ ID NO:191	chromosome 10 open reading frame 76
204161_s_at	0.00013900	12.5	69.3	ENPP4	SEQ ID NO:37	Ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)
205455_at	0.00014640	20.1	333.2	MST1R	SEQ ID NO:42	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
221432_s_at	0.00014780	108.4	34.4	SLC25A28	SEQ ID NO:102	solute carrier family 25, member 28 /// solute carrier family 25, member 28
203082_at	0.00015630	1316.0	435.4	BMS1L	SEQ ID NO:20	BMS1-like, ribosome assembly protein (yeast)
223192_at	0.00015890	391.2	207.2	SLC25A28	SEQ ID NO:116	solute carrier family 25, member 28
226084_at	0.00017240	1660.7	87.5	MAP1B	SEQ ID NO:131	microtubule-associated protein 1B
229587_at	0.00017530	247.0	86.2	UBA2	SEQ ID NO:153	SUMO-1 activating enzyme subunit 2
211071_s_at	0.00018080	2398.5	76.5	AF1Q	SEQ ID NO:60	ALL1-fused gene from chromosome 1q /// ALL1-fused gene from chromosome 1q
214448_x_at	0.00018290	310.0	123.8	NFKBIB	SEQ ID NO:74	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta
225413_at	0.00018660	8130.9	4324.6	USMG5	SEQ ID NO:125	upregulated during skeletal muscle growth 5
235036_at	0.00018930	262.2	19.4	MGC46719	SEQ ID NO:165	hypothetical protein MGC46719
203441_s_at	0.00019180	684.0	72.1	CDH2	SEQ ID NO:31	cadherin 2, type 1, N-cadherin (neuronal)
235247_at	0.00019200	6.2	262.8		SEQ ID NO:167	

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
225096_at	0.00019610	1755.7	703.7	HSA272196	SEQ ID NO:124	hypothetical protein, clone 2746033
205617_at	0.00019960	9.2	23.1	PRRG2	SEQ ID NO:44	proline rich Gla (G-carboxyglutamic acid) 2
225822_at	0.00020110	10.3	468.3	MGC17299	SEQ ID NO:129	hypothetical protein MGC17299
239077_at	0.00020310	146.8	49.3	GALNACT-2	SEQ ID NO:179	chondroitin sulfate GalNAcT-2
218779_x_at	0.00021870	72.0	404.0	EPS8L1	SEQ ID NO:86	EPS8-like 1
50314_i_at	0.00022630	830.5	279.4	C20orf27	SEQ ID NO:190	chromosome 20 open reading frame 27
218792_s_at	0.00023140	74.9	468.6	BSPRY	SEQ ID NO:87	B-box and SPRY domain containing
222664_at	0.00024210	624.9	42.5	KCTD15	SEQ ID NO:109	potassium channel tetramerisation domain containing 15
201869_s_at	0.00024250	290.8	70.5	TBL1X	SEQ ID NO:9	transducin (beta)-like 1X-linked
219855_at	0.00024820	233.0	27.6	NUDT11	SEQ ID NO:94	nudix (nucleoside diphosphate linked moiety X)-type motif 11
203236_s_at	0.00025890	81.3	318.7	LGALS9	SEQ ID NO:22	lectin, galactoside-binding, soluble, 9 (galectin 9)
202167_s_at	0.00026530	770.6	340.7	MMS19L	SEQ ID NO:10	MMS19-like (MET18 homolog, S. cerevisiae)
229223_at	0.00026990	21.7	130.8		SEQ ID NO:152	
201157_s_at	0.00027160	2272.3	1323.6	NMT1	SEQ ID NO:4	N-myristoyltransferase 1
226187_at	0.00027300	32.2	301.2	CDS1	SEQ ID NO:132	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1
239671_at	0.00028050	12.2	43.6		SEQ ID NO:181	CDNA FLJ31085 fis, clone IMR321000037
222746_s_at	0.00028540	8.7	288.5	BSPRY	SEQ ID NO:111	B-box and SPRY domain containing
219858_s_at	0.00029160	12.3	63.1	FLJ20160	SEQ ID NO:96	FLJ20160 protein
210749_x_at	0.00029280	507.7	2452.9	DDR1	SEQ ID NO:59	discoidin domain receptor family, member 1
211778_s_at	0.00029620	20.3	334.6	ZNF339	SEQ ID NO:61	zinc finger protein 339 /// zinc finger protein 339
226876_at	0.00030570	283.5	45.7	MGC45871	SEQ ID NO:141	hypothetical protein MGC45871
230323_s_at	0.00033140	17.4	295.5	LOC120224	SEQ ID NO:157	hypothetical protein BC016153
221665_s_at	0.00033480	20.5	172.5	EPS8L1	SEQ ID NO:105	EPS8-like 1
1007_s_at	0.00033840	469.2	2729.2	DDR1	SEQ ID NO:1	discoidin domain receptor family, member 1
218891_at	0.00034090	218.3	108.6	C10orf76	SEQ ID NO:88	chromosome 10 open reading frame 76
218960_at	0.00034100	25.7	408.5	TMPRSS4	SEQ ID NO:89	transmembrane protease, serine 4
222668_at	0.00034910	573.0	38.2	KCTD15	SEQ ID NO:110	potassium channel tetramerisation domain containing 15
217496_s_at	0.00036040	593.8	172.2	IDE	SEQ ID NO:81	insulin-degrading enzyme
226213_at	0.00036180	27.4	1639.9	ERBB3	SEQ ID NO:133	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
235202_x_at	0.00036460	59.3	14.9	IKIP	SEQ ID NO:166	IKK interacting protein
212736_at	0.00036600	290.0	27.4	BC008967	SEQ ID NO:69	hypothetical gene BC008967
203327_at	0.00036980	410.7	105.9	IDE	SEQ ID NO:26	insulin-degrading enzyme
202597_at	0.00037880	5.1	129.6	IRF6	SEQ ID NO:18	interferon regulatory factor 6
228865_at	0.00037970	9.2	322.3	SARG	SEQ ID NO:149	specifically androgen-regulated protein
205709_s_at	0.00038120	13.4	254.3	CDS1	SEQ ID NO:45	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1
224946_s_at	0.00039420	329.1	681.4	MGC12981	SEQ ID NO:123	hypothetical protein MGC12981
204856_at	0.00039710	80.7	400.7	B3GNT3	SEQ ID NO:39	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3
203317_at	0.00039900	58.0	171.0	PSD4	SEQ ID NO:25	pleckstrin and Sec7 domain containing 4
221958_s_at	0.00040170	171.2	468.6	FLJ23091	SEQ ID NO:106	putative NFkB activating protein 373
201130_s_at	0.00040570	15.3	1183.0	CDH1	SEQ ID NO:3	cadherin 1, type 1, E-cadherin (epithelial)
205458_at	0.00042200	109.4	57.6	MC1R	SEQ ID NO:43	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)
205847_at	0.00042390	71.8	206.0	PRSS22	SEQ ID NO:47	protease, serine, 22
202340_x_at	0.00043030	336.4	72.7	NR4A1	SEQ ID NO:14	nuclear receptor subfamily 4, group A, member 1
215146_s_at	0.00043080	165.6	48.8	KIAA1043	SEQ ID NO:76	KIAA1043 protein
223032_x_at	0.00043320	5068.6	2903.7	PX19	SEQ ID NO:115	px19-like protein
226535_at	0.00044520	15.3	862.3	ITGB6	SEQ ID NO:137	integrin, beta 6
65517_at	0.00045130	50.8	387.0	AP1M2	SEQ ID NO:192	adaptor-related protein complex 1, mu 2 subunit
91826_at	0.00045430	59.7	373.3	EPS8L1	SEQ ID NO:194	EPS8-like 1
238673_at	0.00045640	44.3	578.2		SEQ ID NO:176	
221610_s_at	0.00046860	83.5	569.8	STAP2	SEQ ID NO:104	signal-transducing adaptor protein-2
203779_s_at	0.00047400	17.8	143.2	EVA1	SEQ ID NO:33	epithelial V-like antigen 1
230312_at	0.00047560	91.2	11.6		SEQ ID NO:156	
211855_s_at	0.00047620	355.5	97.2	SLC25A14	SEQ ID NO:62	solute carrier family 25 (mitochondrial carrier, brain), member 14
222830_at	0.00047770	31.3	586.6	TFCP2L2	SEQ ID NO:112	transcription factor CP2-like 2
203780_at	0.00047790	33.5	647.3	EVA1	SEQ ID NO:34	epithelial V-like antigen 1
223233_s_at	0.00048700	37.9	541.0	CGN	SEQ ID NO:117	cingulin
219412_at	0.00049410	6.2	241.9	RAB38	SEQ ID NO:92	RAB38, member RAS oncogene family
219936_s_at	0.00049770	5.8	171.1	GPR87	SEQ ID NO:97	G protein-coupled receptor 87
226226_at	0.00049820	31.5	465.5	LOC120224	SEQ ID NO:134	hypothetical protein BC016153

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
222280_at	0.00050070	312.5	152.0		SEQ ID NO:107	CDNA clone IMAGE:6602785, partial cds
225911_at	0.00050990	6.9	142.2	LOC255743	SEQ ID NO:130	hypothetical protein LOC255743
223295_s_at	0.00053580	463.2	264.9	LUC7L	SEQ ID NO:118	LUC7-like (S. cerevisiae)
212120_at	0.00053760	1118.9	381.7	RHOQ	SEQ ID NO:65	ras homolog gene family, member Q
226584_s_at	0.00053900	81.8	186.8	C20orf55	SEQ ID NO:138	chromosome 20 open reading frame 55
202328_s_at	0.00054270	307.4	127.3	PKD1	SEQ ID NO:13	polycystic kidney disease 1 (autosomal dominant)
208779_x_at	0.00054830	489.8	2385.8	DDR1	SEQ ID NO:55	discoidin domain receptor family, member 1
203783_x_at	0.00055660	33.6	14.8	POLRMT	SEQ ID NO:35	polymerase (RNA) mitochondrial (DNA directed)
208084_at	0.00055660	29.0	347.8	ITGB6	SEQ ID NO:52	integrin, beta 6
213262_at	0.00056350	597.1	48.5	SACS	SEQ ID NO:72	spastic ataxia of Charlevoix-Saguenay (sacsin)
225793_at	0.00058010	1662.4	133.4	MGC46719	SEQ ID NO:128	hypothetical protein MGC46719
226678_at	0.00058120	63.1	171.9	UNC13D	SEQ ID NO:139	unc-13 homolog D (C. elegans)
216949_s_at	0.00058240	83.3	27.2	PKD1	SEQ ID NO:80	polycystic kidney disease 1 (autosomal dominant)
212338_at	0.00058710	28.0	335.5	MYO1D	SEQ ID NO:67	myosin ID
241455_at	0.00059440	7.3	68.8		SEQ ID NO:183	
214577_at	0.00062040	279.3	58.3	MAP1B	SEQ ID NO:75	microtubule-associated protein 1B
220178_at	0.00062110	193.7	48.8	C19orf28	SEQ ID NO:99	chromosome 19 open reading frame 28
201868_s_at	0.00062220	103.1	21.6	TBL1X	SEQ ID NO:8	transducin (beta)-like 1X-linked
201679_at	0.00063150	451.3	212.9	ARS2	SEQ ID NO:6	arsenate resistance protein ARS2
206043_s_at	0.00063910	8.0	67.9	KIAA0703	SEQ ID NO:48	KIAA0703 gene product
226706_at	0.00063930	81.4	847.1	FLJ23867	SEQ ID NO:140	hypothetical protein FLJ23867
210255_at	0.00064190	8.8	36.1	RAD51L1	SEQ ID NO:58	RAD51-like 1 (S. cerevisiae)
208968_s_at	0.00066500	2065.0	1181.4	CIAPIN1	SEQ ID NO:56	cytokine induced apoptosis inhibitor 1
207627_s_at	0.00068160	401.7	205.1	TFCP2	SEQ ID NO:50	transcription factor CP2
203407_at	0.00068500	39.6	1680.0	PPL	SEQ ID NO:29	periplakin
217791_s_at	0.00069580	1777.8	837.7	ALDH18A1	SEQ ID NO:82	aldehyde dehydrogenase 18 family, member A1
225582_at	0.00069740	415.9	44.7	KIAA1754	SEQ ID NO:126	KIAA1754
231721_at	0.00070410	37.7	4.4	JAM3	SEQ ID NO:158	junctional adhesion molecule 3
222859_s_at	0.00072460	24.0	133.1	DAPP1	SEQ ID NO:114	dual adaptor of phosphotyrosine and 3-phosphoinositides
208595_s_at	0.00074160	263.9	122.8	MBD1	SEQ ID NO:54	methyl-CpG binding domain protein 1
212015_x_at	0.00075720	5744.3	3435.4	PTBP1	SEQ ID NO:63	polypyrimidine tract binding protein 1
219856_at	0.00075780	13.9	230.4	SARG	SEQ ID NO:95	specifically androgen-regulated protein
38766_at	0.00075940	85.9	281.7	SRCAP	SEQ ID NO:189	Snf2-related CBP activator protein

Probe set	parametric p-value	mean intensity (resistant)	mean intensity (sensitive)	Gene symbol	Sequence Identifier	Description
P204744_s_at	0.00076150	7537.7	3827.7	IARS	SEQ ID NO:38	isoleucine-tRNA synthetase
239196_at	0.00076210	30.5	550.5	ANKRD22	SEQ ID NO:180	ankyrin repeat domain 22
203718_at	0.00076760	424.0	138.4	NTE	SEQ ID NO:32	neuropathy target esterase
232149_s_at	0.00076810	414.2	127.6	NSMAF	SEQ ID NO:161	neutral sphingomyelinase (N-SMase) activation associated factor
202264_s_at	0.00076920	1513.7	830.7	TOMM40	SEQ ID NO:11	translocase of outer mitochondrial membrane 40 homolog (yeast)
32069_at	0.00077000	147.8	266.2	N4BP1	SEQ ID NO:187	Nedd4 binding protein 1
216862_s_at	0.00078160	901.3	359.6	MTCP1	SEQ ID NO:79	mature T-cell proliferation 1
220370_s_at	0.00079540	306.1	60.5	USP36	SEQ ID NO:101	ubiquitin specific protease 36
242191_at	0.00080180	152.0	35.5		SEQ ID NO:185	LOC400781
203109_at	0.00081840	2445.5	1097.7	UBE2M	SEQ ID NO:21	ubiquitin-conjugating enzyme E2M (UBC12 homolog, yeast)
205780_at	0.00083050	39.8	941.1	CDH2	SEQ ID NO:46	cadherin 2, type 1, N-cadherin (neuronal)
203440_at	0.00\083250	503.5	78.6		SEQ ID NO:30	
238513_at	0.00083510	73.6	618.6	TMG4	SEQ ID NO:175	transmembrane gamma-carboxyglutamic acid protein 4
221550_at	0.00083680	414.1	200.9	COX15	SEQ ID NO:103	COX15 homolog, cytochrome c oxidase assembly protein (yeast)
229030_at	0.00084650	5.9	70.1		SEQ ID NO:151	
226400_at	0.00088590	2284.5	4256.7		SEQ ID NO:135	
37966_at	0.00090730	127.8	9.3	PARVB	SEQ ID NO:188	parvin, beta
212424_at	0.00092430	381.6	115.2	PDCD11	SEQ ID NO:68	programmed cell death 11
228441_s_at	0.00093570	12.0	49.8		SEQ ID NO:147	
203328_x_at	0.00095810	411.3	112.2	IDE	SEQ ID NO:27	insulin-degrading enzyme
201680_x_at	0.00095980	1383.3	765.5	ARS2	SEQ ID NO:7	arsenate resistance protein ARS2
243302_at	0.00096750	14.2	29.1		SEQ ID NO:186	
219969_at	0.00097320	102.8	21.4	CXorf15	SEQ ID NO:98	chromosome X open reading frame 15
212016_s_at	0.00099210	4187.6	2276.0	PTBP1	SEQ ID NO:64	polypyrimidine tract binding protein 1

Example 2

The following example describes the identification and further investigation of a target gene identified using the gene expression profile disclosed herein.

In this experiment, the present inventors describe research to examine the influence of E-cadherin-regulatory molecules on non-small cell lung cancer (NSCLC) response to EGF receptor (EGFR) inhibitors.

The EGFR, a member of the erbB family of tyrosine kinases (erbB1-4) plays a major role in transmitting stimuli that lead to NSCLC cellular proliferation and survival. EGFR, highly expressed in NSCLC, is a primary target for NSCLC therapeutic intervention. In clinical trials, 11-20% of patients with chemo-refractory advanced stage NSCLC responded to treatment with EGFR inhibitors such as gefitinib (Iressa®, ZD1839). Currently, there are no markers that predict which patients will respond to treatment. NSCLC patients with poor survival have decreased expression of E-cadherin, a cell adhesion molecule. E-cadherin expression is regulated by the wnt pathway and by zinc finger transcription factors including δEF1/ZEB1 and SIP1/ZEB2. Higher levels of protein expression of E-cadherin were detected in gefitinib sensitive NSCLC cell lines and expression was absent in gefitinib resistant lines. Conversely, expression of the E-cadherin inhibitors ZEB1 and SIP1 was higher in gefitinib resistant cell lines. The hypothesis of this project is that expression of E-cadherin and its regulatory molecules predict response to EGFR inhibitors, and modulating E-cadherin regulatory proteins may augment response to EGFR inhibitors in non-small cell lung cancer.

E-cadherin, a calcium-dependent epithelial cell adhesion molecule, plays an important role in tumor invasiveness and metastatic potential. Reduced E-cadherin expression is associated with tumor cell dedifferentiation, advanced stage and reduced survival in patients with NSCLC. At the transcriptional level, the wnt/β-catenin signaling pathway regulates DE-cadherin expression. The present inventors have reported that inhibition of GSK3β, involved in the proteasomal degradation of β-catenin, lead E-cadherin upregulation (Fig. 2). E-cadherin transcription is also regulated by zinc finger transcription factors including, Snail, Slug, ZEB1 and SIP1. They repress E-cadherin expression by binding to its promoter and recruiting HDAC (Fig. 2). The inventors have reported that inhibiting the ZEB1 or HDAC expression lead to upregulation of E-cadherin in NSCLC cell lines.

In this experiment, the inventors used NSCLC cell lines to: (1) evaluate the growth inhibitory properties of EGFR inhibitors by MTT assays, (2) to identify molecular molecules through DNA microarrays and westerns that predict response to EGFR inhibitors and (3) to design combination therapies that enhance the effect of the EGFR inhibitors. Cell lines were screened for expression of members of the EGFR and Wnt signaling pathways. E-cadherin expression was found to be lacking in gefitinib resistant cell lines and activated in gefitinib sensitive lines. Therefore, the expression of zinc finger transcription factors involved in E-cadherin repression was investigated. It was determined that gefitinib resistant lines have high ZEB1 and/or SIP1 expression, and expression is lacking in gefitinib-sensitive lines.

The inventors proposed that SIP1 and ZEB1 expression predicts EGFR tyrosine kinase inhibitors resistance and that modulating the molecular mechanism that regulate E-cadherin expression will enhance sensitivity to EGFR inhibitors. The proposal will be tested by manipulating E-cadherin expression and measuring the effect on response to gefitinib. Results of this work will be evaluated in clinical trials in patients with NSCLC.

Results

EGFR, pEGFR, Her2, ErbB3 and Erb4 expression in NSCLC:

EGFR, Her-2 and ErbB3 cell surface expression was evaluated using flow cytometry (Table 2). The majority of NSCLC cell lines (15/18) had a high percentage of EGFR positive cells and three had low or now EGFR expression. The two BAC cell lines, H322 and H358, had high expression of EGFR and Her2.

Table 2

Cell Line	FACS %EGFR/ MFI	FACS %Her2/ MFI	FACS %ErbB3/ MFI	IC 50 uM ZD 1839
Adenocarcinoma				
Calu3	98%/8.9	100/37	32/4.3	<1
Colo699	0/0	0/0	57/2.3	4.1
H125	100/34	91/2.8	0/0	4.7
H2122	94/5.1	73/4	80/5	4.8
H1435	98/14	ND	94/6.4	7.6
A549	99/14	72/2.4	54/3.5	8.4
H441	78/6.9	79/2.6	0/0	11.7
H1648	98/5.7	78/2.7	0/0	11.5
Bronchoalveolar				
H322	100/16	96.5.5	ND	<1
H358	ND	ND	ND	<1
Squamous Cell				

NE18	100/16	98/3.3	35/5.7	8
H1703	99/15	65/2.6	0/0	9.3
H157	93/13	62/1.8	0/0	10.1
H520	0/0	0/0	0/0	10.3
H1264	100/14	43/1.9	0/0	10.2
Large Cell				
H1334	100/23	74/3.2	99/10	3.8
H460	37/1.9	57/1.4	0/0	9/9

The presence of phosphorylated EGFR (pEGFR) versus EGFR was evaluated by Western blotting in 18 NSCLC cell lines (Fig. 3, shows 15 cell lines). EGFR was detected in the majority of NSCLC cell lines, whereas only a subset of these cell lines had (pEGFR).

Effects of EGFR inhibitors on human lung cancer cells growth:

The growth inhibitory effect of gefitinib, on 18 NSCLC cell lines was evaluated using the MTT assay (Table 2). There was no correlation between the EGFR expression and gefitinib response. The change in pEGFR following gefitinib treatment was evaluated in two sensitive cell lines, H1334 and H322, and two resistant cell lines, H1264 and H1648 (Fig. 4). Gefitinib inhibited the phosphorylated "active" form of EGFR in sensitive cell lines.

Based on the *in vitro* experiments, athymic nude mice bearing human NSCLC xenografts were treated with EGFR inhibitors ZD1839 or C225. Growth delay was evident in tumors after treatment with either agent (Fig. 5).

E-cadherin, SIP1 and ZEB1 in NSCLC cell lines using microarray and RT-PCR and western blotting:

High density oligonucleotide microarray (IOAM) analysis of gene expression levels of selected genes was developed from 11 NSCLC cell lines. These cell lines included 2 gefitinib sensitive lines ($IC_{50}<1\mu M$), 5 gefitinib resistant lines ($IC_{50}\geq10\mu M$), and 4 lines with intermediate sensitivity ($IC_{50}>1\mu M$, $<10\mu M$). The expression of E-cadherin, SIP1 and ZEB 1 was evaluated and compared to their expression in normal bronchial epithelium using the Gene Spring program (Fig. 6).

E-cadherin expression was more pronounced in gefitinib sensitive lines absent in gefitinib resistant lines. This expression pattern was confirmed using western blotting and real time PCR (RT-PCR) (Fig. 7).

As discussed above, regulation of E-cadherin expression involves the zinc finger transcription factors ZEB1 and SIP1. Expression of both transcription factors was evaluated using real time RT-PCR. ZEB 1 and SIP 1 were expressed in the gefitinib resistant lines and absent in the gefitinib sensitive lines (Fig. 8). The expression of Slug, Snail, Wnt7a, β-catenin, γ-catenin, α-catenin and GSK3β was also evaluated using Western blot analysis or RT-PCR. None of theses molecules had a differential pattern of expression in the NSCLC lines (data not shown).

In summary, there was no correlation between gefitinib sensitivity and EGFR expression. E-cadherin was detected preferentially in gefitinib sensitive lines. Conversely, the zinc finger transcription factors, ZEB1 and SIP1, involved in E-cadherin inhibition were expressed in gefitinib resistant lines and absent in gefitinib sensitive lines.

Example 3

This example describes the evaluation of the detrimental effect of the zinc finger proteins ZEB1 and SIP1 on NSCLC cell lines sensitivity to EGFR inhibitors.

In the first part of this experiment, adenoviral constructs containing ZEB1 or SIP1 are used to overexpress these proteins in gefitinib sensitive cell lines. MTT assay will assess changes in gefitinib sensitivity. In the second part of this experiment, stably transfected ZEB1 and SIP1 cell lines and untransfected cell lines are implanted into nude mice. Transplanted mice are treated with gefitinib and the response is compared between the two groups.

Example 4

This example describes the determination of the molecular mechanisms that improve the response to EGFR inhibitors in NSCLC cell lines *in vitro* and *in vivo*.

In the first part of this experiment, the effect of "silencing" the E-cadherin transcriptional repressors, SIP1 and ZEB1, on NSCLC cell lines response to ZD1839 is examined. To directly examine the role of the zinc-finger transcription factors, SIP1 and ZEB1 on gefitinib responsive lines, the effect of siRNA is developed and tested (Fig. 9). siRNA is prepared for different regions of SIP1 and ZEB1 using the silencer kit from Dharmacon (Colorado). Their efficacy is tested by RT-PCR. The most effective siRNA for SIP1 and ZEB1 are then introduced, individually or in combination, into gefitinib resistant lines. The effect of these siRNAs on gefitinib responsiveness is evaluated by MTT assay. ZEB1 antibody (Santa Cruz, California) and SIP1 antibody (a gift from Dr. van Grunsven) are used to evaluate the efficacy of RNA inhibition.

In the second part of this experiment, the effect of inhibiting GSK3β on gefitinib response in NSCLC cell lines is examined. GSK3β phosphorylates β-catenin leading to its ubiquitination and destruction. GSK3β inhibitors, such as lithium, increased E-cadherin expression in NSCLC cell lines. GSK3β function is inhibited with an adenovirus (pAdTrack-CMV) encoding a dominant-negative GSK3β (dnGSK3β). To determine the effectiveness of this dnGSK3 the expression of non-phosphorylated β-catenin and E-cadherin is evaluated by western blot. NSCLC cell lines stably transfected with the dnGSK3β construct are generated. The effect of inhibiting GSK3β on NSCLC cell lines response to gefitinib are evaluated using MTT assays.

In the third part of this experiment, the effect of E-cadherin on gefitinib sensitivity is evaluated. Resistant NSCLC lines are transfected with E-cadherin encoding constructs. Changes in NSCLC cell lines response to gefitinib are assessed by MTT assay. Gefitinib-sensitive lines that express E-cadherin are treated with an E-cadherin antibody (Zymed) and the effect on gefitinib responsiveness assessed by MTT assay. The results determine whether expression of E-cadherin itself is sufficient to determine gefitinib sensitivity, or if sensitivity is a reflexion of events occurring upstream of it.

In the fourth part of this experiment, the effect of gefitinib responsiveness on NSCLC cell lines is augmented *in vivo*. Based on findings from the above *in vitro* experiments, the best treatment that enhances gefitinib sensitivity in NSCLC cell lines is selected for *in vivo* experiments in nude mice. Previously, the inventors showed an inhibitory effect of gefitinib alone on NSCLC xenografts growth (see above). The combination of gefitinib with one of the above-evaluated interventions is tested in athymic nude mice bearing human NSCLC xenografts. E-cadherin inducible cell lines from the *in vitro* experiments are inoculated subcutaneously in nude mice. Mice are treated with gefitinib with and without the agent that improved the gefitinib sensitivity. The two groups are evaluated for differences in tumor growth inhibition. Expression of E-cadherin, SIP1 and ZEB1 are evaluated both prior to and post-treatment by real-time RT-PCR and immunohistochemistry. ZEB1 antibody (Santa Cruz, California) and SIP1 antibody (a gift from Dr. van Grunsven) are used in the immunohistochemistry. However, new antibodies can readily be generated if the above antibodies are not effective at detecting proteins in the IHC assays.

The results of these experiments dissect out the events leading to gefitinib resistance in order to develop treatment modifications that bypass resistance.

Each publication and reference cited herein is incorporated herein by reference in its entirety. U.S. Provisional Application Serial No. 60/538,682, filed January 23, 2004, is incorporated herein by reference in its entirety.

5   While various embodiments of the present invention have been described in detail, it is apparent that modifications and adaptations of those embodiments will occur to those skilled in the art. It is to be expressly understood, however, that such modifications and adaptations are within the scope of the present invention, as set forth in the following claims.